

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/12, C07K 14/47, 16/18, C12Q 1/68, G01N 33/574, 33/68, A61K 38/17</b>		<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/21331</b> <b>(43) International Publication Date:</b> 22 May 1998 (22.05.98)
<b>(21) International Application Number:</b> PCT/US97/20674 <b>(22) International Filing Date:</b> 7 November 1997 (07.11.97) <b>(30) Priority Data:</b> 08/747,547 12 November 1996 (12.11.96) US <b>(71) Applicant (for all designated States except US):</b> INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> AKERBLOM, Ingrid, E. [US/US]; 1234 Johnson Street, Redwood City, CA 94061 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). MURRY, Lynn, E. [US/US]; 1124 Los Trancos Road, Portola Valley, CA 94028 (US). GOLI, Surya, K. [IN/US]; 620 Iris Avenue #338, Sunnyvale, CA 94086 (US). HAWKINS, Phillip, R. [US/US]; 750 N. Shoreline Boulevard #96, Mountain View, CA 94034 (US). <b>(74) Agent:</b> BILLINGS, Lucy, J.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).		<b>(81) Designated States:</b> AT, AU, BR, CA, CH, CN, DE, DK, ES, FI, GB, IL, JP, KR, MX, NO, NZ, RU, SE, SG, US, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
<b>(54) Title:</b> HUMAN BREAST TUMOR-SPECIFIC PROTEINS			
<pre>          9      18      27      36      45      54 5' NOT CCA AAT CAC TCA TCG TTT GTG AAA GCT GAG CTC ACA GCA AAA CAA GCC ACC           63      72      81      90      99      108 ATG AAG CTG TCG GTG TGF CTC CTG GTC ACG CTG GCC CTC TGC TGC TAC CAG M K L S V C L L L V T L A L C C Y Q           117     126     135     144     153     162 GCC AAT GCC GAG TTC TGC CCA GCT CTT GTT TGT GAG CTG TTA GAC TTC TTC TTC A H A K F C P A L V S E L L D F F F           171     180     189     198     207     216 ATT AGT GAA CCT CTG TTC AAG TTA AGT CTT GGC AAA TTT GAT GCC CCT CCG GAA I S E P L F K L S L A K F D A P P E           225     234     243     252     261     270 GCT GTT GCA GCC AAG TTA GGA GTG AAG AGA TGC ACG GAT CAG ATG TCC CTT CAG A V A A K L G V K R C T D Q E S L Q           279     288     297     306     315     324 AAA CGA ACG CTC ATT GCG GAA GTC CTG GTG AAA ATA TTG AAG AAA TGT AGT GTG K R S L I A E V L V K I L K K C S V           333     342     351     360     369     378 TGA CAT GTA AAA ACT TTC ATC CTG GTT TCC ACT GTC TTT CAA TGA CAC CCT GAT           387     396     405 CTT CAC TGC AGA ATG TAA AAG TTT CAA C 3'</pre>			
<b>(57) Abstract.</b>  The present invention provides polynucleotides that identify and encode two human steroid binding proteins (hSBP). The invention provides for genetically engineered expression vectors and host cells comprising the nucleic acid sequences encoding hSBP polypeptides. The invention also provides for the use of substantially purified hSBP polypeptides, antagonists, and nucleotide sequences (e.g., antisense sequences) in pharmaceutical compositions for the treatment of diseases associated with the expression of hSBP, specifically in the treatment of breast cancer. The invention also describes diagnostic assays for the detection of breast cancer in a susceptible or affected patient. The diagnostic assays utilize compositions comprising the polynucleotides encoding hSBP polypeptides or the complements thereof, which hybridize with the genomic sequence or the transcript of polynucleotides encoding hSBP or anti-hSBP antibodies that specifically bind to an hSBP polypeptide.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## HUMAN BREAST TUMOR-SPECIFIC PROTEINS

### TECHNICAL FIELD

The present invention relates to nucleic acid and amino acid sequences of proteins that are differentially expressed in human breast tumor cells and to the use of these sequences in the diagnosis, study, prevention and treatment of disease.

### BACKGROUND ART

Development of breast cancer is associated with multiple genetic changes associated with alterations in expression of specific genes. Breast cancer tissues express genes that are not expressed, or expressed at lower levels, by normal breast tissue. Thus, it is possible to differentiate between normal (non-cancerous) breast tissue and cancerous breast tissue by analyzing differential gene expression between tissues. In addition, there may be several possible alterations that lead to the various possible types of breast cancer. Thus, different types of breast tumors (e.g., invasive vs. non-invasive, ductal vs. axillary lymph node) can be differentiable one from another by the identification of the differences in genes expressed by different types of breast tumor tissues (Porter-Jordan et al. 1994 Hematol Oncol Clin North Am 8:73-100). Breast cancer can thus be generally diagnosed by detection of expression of a gene or genes associated with breast tumor tissue. Where enough information is available about the differential gene expression between various types of breast tumor tissues, the specific type of breast tumor can also be diagnosed.

Nucleotide and amino acid sequences associated with breast tumors can serve as genetic markers of inheritable breast cancer. Genetic changes on chromosome 17 are the most frequently identified events associated with breast tumors. At least four markers on chromosome 17 have been identified: p53 on 17p13.1, regions of loss of heterozygosity (LOH) on 17p13.3 and 17q12-qter, the breast/ovarian cancer locus (BRCA-1) on 17q21, and a fourth breast cancer growth suppressor gene on chromosome 17 (Casey et al. 1993 Hum Molec Genet 2:1921-1927).

Such genetic markers can also be useful in identifying patients susceptible to breast cancer. For example, the genetic marker BRCA-1 has been linked to a susceptibility of developing breast and/or ovarian cancer at a young age in a number of families (Hall et al. 1990 Science 250:1684-1689; Solomon et al. 1991 Cytogenet Cell Genet 58:686-738). The cumulative risks of developing breast cancer associated with the BRCA-1 marker are 50% at 50 years and 82% at 70 years (Easton et al. 1993 Am J Hum Genet 52:678-701). However, since the gene encoding BRCA-1 has not been cloned or sequenced, identification of an individual carrier

of BRCA-1 is not possible without use of linkage analysis. Linkage analysis is generally not feasible in clinical practice since the genetic epidemiology required is tedious, if not impossible, in most cases (Kent et al. 1995 Europ J Surg Oncol 21:240-241).

5 The discovery of nucleotide sequences and polypeptides encoding proteins associated with breast cancer would satisfy a need in the art by providing new means of diagnosing and treating breast cancer.

### DISCLOSURE OF THE INVENTION

10 The present invention features two human steroid binding proteins (hereinafter referred to individually as hSBP1, and hSBP2, and collectively as hSBP), and the full-length nucleotide sequences encoding these proteins, which are differentially expressed in human breast tumor tissue. The transcripts encoding these proteins are present in breast tumor tissue. The first polypeptide, referred to hereinafter as human steroid binding protein C1 (hSBP1), is characterized as having amino acid sequence homology to rat prostatic binding proteins C1 and C2 (PSC1\_RAT and PSC2\_RAT' respectively) and nucleotide sequence homology to hamster  
15 FHG 22 (GI 206441). The second polypeptide, referred to hereinafter as human steroid binding protein C2 (hSBP2), is characterized as having identity to human mammaglobin and homology to rat prostatic binding protein C3 (GI 206448). Accordingly, the invention features two substantially purified human steroid binding proteins, as shown in amino acid sequences of SEQ ID NO:1 and SEQ ID NO:3.

20 One aspect of the invention features isolated and substantially purified polynucleotides that encode hSBP. In a particular aspect, the polynucleotide is the nucleotide sequence of SEQ ID NO:2 and SEQ ID NO:4. In addition, the invention features polynucleotide sequences that hybridize under stringent conditions to SEQ ID NO:2 and SEQ ID NO:4.

25 The invention additionally features nucleic acid sequences encoding hSBP polypeptides, oligonucleotides, peptide nucleic acids (PNA), fragments, portions or antisense molecules thereof, and expression vectors and host cells comprising polynucleotides that encode hSBP. The present invention also relates to antibodies which bind specifically to an hSBP polypeptide, pharmaceutical compositions comprising substantially purified hSBP, fragments thereof, or antagonists of hSBP, in conjunction with a suitable pharmaceutical carrier, and methods for  
30 producing hSBP.

### BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ

ID NO:2) of human steroid binding protein C1, hSBP1. The alignment was produced using MacDNAsis software (Hitachi Software Engineering Co Ltd, San Bruno, CA).

Figures 2A and 2B shows the amino acid sequence (SEQ ID NO:3) and nucleic acid sequence (SEQ ID NO:4) of human steroid binding protein C2, hSBP2 (MacDNAsis software, Hitachi Software Engineering Co Ltd).

Figure 3 shows the northern analysis for the consensus sequence (SEQ ID NO:2) for hSBP1 (Incyte clone 606491). The northern analysis was produced electronically using LIFESEQ™ database (Incyte Pharmaceuticals, Palo Alto CA). The abundance data (Abun) represent the number of transcripts of the gene of interest in the cDNA library. Percent abundance is calculated by dividing the number of transcripts of a gene of interest present in a cDNA library by the total number of transcripts in the cDNA library.

Figure 4 shows the northern analysis for the consensus sequence (SEQ ID NO:4) (LIFESEQ™ database. Incyte Pharmaceuticals, Palo Alto CA).

Figure 5 shows the amino acid sequence alignments among hSBP1 (606491; SEQ ID NO:1) rat prostatic binding proteins C1 and C2 (SEQ ID NOS:5 and 8), and rabbit uteroglobin (SEQ ID NO:9), produced using the multisequence alignment program of DNASTar software (DNASTar Inc. Madison WI).

Figure 6 shows the amino acid sequence alignments among hSBP2 (SEQ ID NO:3) human mammaglobin (GI 1199595; SEQ ID NO:10), and rat prostatic binding protein C3 (GI 206453; SEQ ID NO:12), produced using the multisequence alignment program of DNASTar software (DNASTar Inc, Madison WI).

Figures 7A and 7B shows the nucleotide sequence alignments between hSBP1 (606491; SEQ ID NO:2), hamster FHG22 (GI 1045204; SEQ ID NO:7), and rat prostatic binding protein C1 (GI 206441; SEQ ID NO:6).

Figures 8A and 8B show the nucleotide sequence alignments between hSBP2 (602516; SEQ ID NO:4), human mammaglobin (GI 1199595; SEQ ID NO:11), and rat prostatic binding protein C3 (GI 206452; SEQ ID NO:13).

## MODES FOR CARRYING OUT THE INVENTION

### Definitions

"Nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which can be single- or double-stranded, and represent the sense or antisense strand.

Similarly, "amino acid sequence" as used herein refers to an oligopeptide, peptide, polypeptide, or protein sequence. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally-occurring protein molecule, "amino acid sequence" and like terms (e.g., polypeptide, or protein) are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

"Peptide nucleic acid" as used herein refers to a molecule which comprises an oligomer to which an amino acid residue, such as lysine, and an amino group have been added. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary (template) strand of nucleic acid (Nielsen PE et al (1993) Anticancer Drug Des 8:53-63).

As used herein, "SBP" refers to the amino acid sequences of substantially purified steroid binding protein obtained from any species, particularly mammalian, including bovine, ovine, porcine, murine, equine, and preferably human, from any source whether natural, synthetic, semi-synthetic or recombinant. The term "hSBP" as used herein refers to human steroid binding protein and is meant to encompass hSBP1 and hSBP2 polypeptides collectively.

As used herein, "antigenic amino acid sequence" means an amino acid sequence that, either alone or in association with a carrier molecule, can elicit an antibody response in a mammal.

A "variant" of hSBP is defined as an amino acid sequence that is altered by one or more amino acids. The variant can have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant can have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Similar minor variations can also include amino acid deletions or insertions, or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological or immunological activity can be found using computer programs well known in the art, for example, DNASTar software.

A "deletion" is defined as a change in either amino acid or nucleotide sequence in which one or more amino acid or nucleotide residues, respectively, are absent.

An "insertion" or "addition" is that change in an amino acid or nucleotide sequence which has resulted in the addition of one or more amino acid or nucleotide residues, respectively, as compared to the naturally occurring hSBP.

A "substitution" results from the replacement of one or more amino acids or nucleotides

by different amino acids or nucleotides, respectively.

The term "biologically active" refers to a hSBP having structural, regulatory, or biochemical functions of a naturally occurring hSBP. Likewise, "immunologically active" defines the capability of the natural, recombinant or synthetic hSBP, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "derivative" as used herein refers to the chemical modification of a nucleic acid encoding hSBP or the encoded hSBP. Illustrative of such modifications would be replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative would encode a polypeptide which retains essential biological characteristics of natural hSBP.

As used herein, the term "substantially purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated.

"Stringency" typically occurs in a range from about  $T_m - 5^\circ\text{C}$  ( $5^\circ\text{C}$  below the  $T_m$  of the probe) to about  $20^\circ\text{C}$  to  $25^\circ\text{C}$  below  $T_m$ . As will be understood by those of skill in the art, a stringency hybridization can be used to identify or detect identical polynucleotide sequences or to identify or detect similar or related polynucleotide sequences.

The term "hybridization" as used herein shall include "any process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs J (1994) Dictionary of Biotechnology, Stockton Press, New York NY). Amplification as carried out in the polymerase chain reaction technologies is described in Dieffenbach CW and GS Dveksler (1995, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY).

#### **Preferred Embodiments**

The present invention relates to hSBP and to the use of hSBP nucleic acid and amino acid sequences in the study, diagnosis, prevention and treatment of disease. cDNAs encoding a portion of hSBP were predominantly found in cDNA libraries derived from breast tumor tissue (Figures 3 and 4). The abundance data (Abun) reflects the relative level of expression the hSBP sequence in the breast, thymus and prostatic cDNA libraries, with the percentage abundance (Pct Abun) representing the percent of total expressed mRNAs that are homologous to the hSBP sequence.

The present invention also encompasses hSBP variants. A preferred hSBP variant is one

having at least 80% amino acid sequence similarity to an amino acid sequence of an hSBP (i.e., an hSBP1 amino acid sequence (SEQ ID NO:1) or an hSBP2 amino acid sequence (SEQ ID NO:3). A more preferred hSBP variant is one having at least 90% amino acid sequence similarity to SEQ ID NO:1 or SEQ ID NO:3. A most preferred hSBP variant is one having at least 95% amino acid sequence similarity to SEQ ID NO:1 or SEQ ID NO:3.

Nucleic acids encoding the human hSBP of the present invention were first identified in cDNA, Incyte Clones 606491 and 602615 from breast tumor cell cDNA library BRSTTUT01 through a computer-generated search for amino acid sequence alignments. A consensus sequence for each of hSBP1 (SEQ ID NO:2) and hSBP2 (SEQ ID NO: 4) was derived from the overlapping and/or extended nucleic acid sequences as shown in the tables below.

Table 1. Clones from which the consensus sequence (SEQ ID NO:2) of hSBP-C1 was derived.

Sequence I.D.	cDNA Library	Sequence I.D.	cDNA Library	Sequence I.D.	cDNA Library
419412H1	BRSTNOT01	606371H1	BRSTTUT01	1212741H1	BRSTTUT01
603148H1	BRSTTUT01	606491H1	BRSTTUT01	1215122H1	BRSTTUT01
603224H1	BRSTTUT01	825519H1	PROSNOT06	1216374H1	BRSTTUT01
604290H1	BRSTTUT01	967077H1	BRSTNOT05	1217152H1	BRSTTUT01
604954H1	BRSTTUT01	1209955H1	BRSTNOT02		
605120H1	BRSTTUT01	1212005H1	BRSTTUT01		



Table 2. Clones from which the consensus sequence (SEQ ID NO:4) of hSBP-C2 was derived.

Sequence I.D.	cDNA Library	Sequence I.D.	cDNA Library	Sequence I.D.	cDNA Library
410758H1	BRSTNOT01	899784	BRSTTUT03	968163H1	BRSTNOT05
419059H1	BRSTNOT01			977969H1	BRSTNOT02
598065H1	BRSTNOT02	899895H1	BRSTTUT03	10000571H1	BRSTNOT03
601000H1	BRSTNOT02	900118H1	BRSTTUT03	1002776H1	BRSTNOT03
602615H1	BRSTTUT01	901009H1	BRSTTUT03	1004904H1	BRSTNOT02
603548H1	BRSTTUT01	902666H1	BRSTTUT03	1210748H1	BRSTNOT02
603234H1	BRSTTUT01	902354H1	BRSTTUT03		
603999H1	BRSTTUT01	959213H1	BRSTTUT03	1212473H1	BRSTTUT01
605093H1	BRSTTUT01	959506H1	BRSTTUT03	1213350H1	BRSTTUT01
605204H1	BRSTTUT01	960045H1	BRSTTUT03	1213570H1	BRSTTUT01
605215H1	BRSTTUT01	960118H1	BRSTTUT03	1213702H1	BRSTTUT01
605561H1	BRSTTUT01	960656H1	BRSTTUT03	1214253H1	BRSTTUT01
606191H1	BRSTTUT01	962153H1	BRSTTUT03	1214304H1	BRSTTUT01
606289H1	BRSTTUT01	962283H1	BRSTTUT03	1214401H1	BRSTTUT01
606611H1	BRSTTUT01	962488H1	BRSTTUT03	1215366H1	BRSTTUT01
606664H1	BRSTTUT01	962656H1	BRSTTUT03	1215626H1	BRSTTUT01
607089H1	BRSTTUT01	962907H1	BRSTTUT03	1216546H1	BRSTTUT01
897552H1	BRSTNOT05	963043H1	BRSTTUT03	1216653H1	BRSTTUT01
898516H1	BRSTTUT03	963046H1	BRSTTUT03	1216659H1	BRSTTUT01
898821H1	BRSTTUT03	964108H1	BRSTTUT03	1216778H1	BRSTTUT01
899628H1	BRSTTUT03	968127H1	BRSTNOT05		

- 25 The nucleic acid sequence of SEQ ID NO:2 encodes the hSBP1 amino acid sequence, SEQ ID NO:1. The nucleic acid sequence of SEQ ID NO:4 encodes the hSBP2 amino acid sequence. SEQ ID NO:3.

The present invention is based, in part, on the chemical and structural homology between:

1) The amino acid sequence of hSBP1 and rat prostatic binding protein C1 (GI 206442; Delaey et al. 1983 Eur J Biochem 133:645-649) rat prostatic binding protein C2 (Delaey et al. 1987 Nucl Acid Res 15:1627-1641 and rabbit uteroglobin (Menne et al. 1982 Proc Natl Acad Sci USA 79:4853-4857; Figure 5) and the amino acid sequences of hSBP2, human mammaglobin (GI 1100595; SEQ ID NO:10) and rat prostatic binding protein C3 (GI 206453; SEQ ID NO:12;

Figure 6); and 2) The nucleotide sequence encoding hSBP1, rat prostatic binding protein C1 (GI 206442; Delaey et al. supra), and hamster FHG22 (GI 1045204; Dominguez 1995 FEBS Letters 376:257-263; Figures 7A and 7B); and hSBP2, human mammaglobin (GI 1199595; Watson et al. 1996 Cancer Res 56:860-865), and rat prostatic binding protein C3 (GI 206452; Parker et al. 1983 J Biol Chem 258:12-15) (Figures 8A and 8B).

Rat prostatic binding protein (rPBP) is a tetrameric, steroid-binding glycoprotein found in rat ventral prostate, and is the principal protein in rat prostatic fluid (Delaey et al. supra; Parker et al. supra; Heyns et al. 1977 Eur J Biochem 78:221-230; Heyns et al. 1977 Biochem Biophys Res Commun 77:1492-1499; Parker et al. 1978 Eur J Biochem 85:399-406). The rPBP tetramer is composed of two subunits: one subunit containing the polypeptides C1 and C3; and the other subunit containing the polypeptides C2 and C3 (Heyns et al. 1978 Eur J Biochem 89:181-186). rPBP C3 is homologous to human mammaglobin, which in turn is homologous to human Clara cell 10-kilodalton protein and rabbit uteroglobin (Watson et al. supra).

Although rat PBP is primarily expressed in the testes (Lindzey et al. 1994 Vitamins Hormones 49:383-32), transgenic animals harboring a construct containing the 5' flanking region of the rat PBP-C3 gene linked to the coding region for the simian virus 40 large tumor antigen express the transgene in both the prostate and the mammary gland (Allison et al. 1989 Mol Cell Biol 9(5): 2254-2257). The expression of the C3 transgene varies with the sex of the transgenic animal; male transgenic animals express the rat PBP C3 transgene in the prostate and develop prostate carcinoma, while the females express the transgene in the mammary gland and develop atypical mammary hyperplasia (Maroulakou et al. 1994 Proc Natl Acad Sci USA 91:11236-40). Expression of rPBP is regulated by androgenic steroid (e.g., testosterone) partly by stimulating rates of transcription and partly by effects on RNA stability (Parker et al. 1977 Cell 12:401-407; Heyns et al. 1977 Biochem Biophys Res Commun 77:1492-1499; Parker et al. 1979 Proc Natl Acad Sci USA 76:1580-1584; Page et al. 1982 Mol Cell Endocr 27:343-355).

rPBP is similar to estramucine binding protein (EMBP) (Heyns et al. 1977 Eur J Biochem 78:221-30). EMBP is a 46-kDa heterodimer consisting of two closely related subunits, which upon reductive cleavage of disulfide bridges, each subunit is divided into two components. The subunits differ with respect to the components C1 and C2, but share C3 (Bjork et al. 1995 The Prostate (1995) 27:70-83). EMBP binds estramucine (Appelgren et al. 1979 Acta Pharmacol Toxicol 43:368-74; Försgren et al. 1979 Cancer Res 39:5155-64; Høisaeter et al. 1981 J Steroid Biochem 14: 251-60), but does not bind free estrogens (Høisaeter et al. 1981 J Steroid Biochem

14:251-260; Försgren et al. 1979 Proc Natl Acad Sci USA 76:3149-3150). Estramucine, a nitrogen mustard derivative of 17 $\beta$ -estradiol (Mittelman et al. 1977 Cancer Treat Rep 61:307-10; Johnson et al. 1971 Scand J Urol Nephrol 5:103-7), is used to treat patients with prostatic carcinoma. Expression of EMBP is androgen-regulated; this androgen-dependency of EMBP  
5 tends to decline with the transformation of prostatic tissue into biologically more malignant disease (Shiina et al. 1996 Brit J Urol 77:96-101). The ratio of EMBP to dihydroxytestosterone is an indicator of the malignant potential of prostatic carcinoma (Shiina et al. supra).

Rabbit uteroglobin, a homodimeric protein coupled by two disulfide linkages, binds progesterone and structurally related steroids, is also a substrate for transglutaminases, inhibits  
10 phospholipase A<sub>2</sub> activity, and may interfere with the immune and inflammatory activity of several cell types (Miele et al. 1994 J Endocrinol Invest 17:679-692; Miele et al. 1987 Endocrinol Rev 8:474-490). Expression of uteroglobin is regulated by tissue-specific response to steroid hormones (Sandmoller et al. 1994 Oncogene 9:2805-2815).

FHG22 protein was isolated from a female minus male subtracted cDNA library obtained  
15 from the sexually dimorphic Syrian hamster Harderian glands (Dominguez supra). FHG nucleotide and amino acid sequence are similar to the subunits from rat prostatic steroid binding protein C1, uteroglobin (Miele et al. 1994 J Endocrinol Invest 17:679-692), major cat allergen Fel dI (chain I), and mouse salivary androgen binding proteins (subunit  $\alpha$ ) (Karn et al. 1993 Biochem Genet 32:271-277; Dominguez supra). Expression of FHG22 is tissue and sex-  
20 dependent (Dominguez supra).

hSBP1 and rat prostatic binding protein C1 share 55% nucleotide sequence identity at the nucleotide sequence level, whereas hSBP1 and hamster FHG22 share 72% nucleotide sequence identity. hSBP1 is 90 amino acids in length; the amino acid sequence of hSBP1 has 49% identity  
25 with the amino acid sequence of rat prostatic binding protein C1 (SEQ ID NO:5), 44% identity with the amino acid sequence of rat prostatic binding protein C2 (SEQ ID NO:8), and 28% identity with the amino acid sequence of rabbit uteroglobin (SEQ ID NO:9) (Figure 5).

hSBP2 is 93 amino acids in length and shares 99% nucleotide sequence identity with human mammaglobin; the nucleotide sequence of hSBP2 is about 43% identical to the nucleotide sequence of rat prostatic binding protein C3 (Figures 8A and 8B). The amino acid sequence of  
30 hSBP2 is 62% identical to the amino acid sequence of rat prostatic protein C3, and 100% identical to the amino acid sequence of human mammaglobin (Figure 6). Thus, hSBP-C3 is identical to human mammaglobin.

### The hSBP Coding Sequences

The nucleic acid and deduced amino acid sequences of hSBP are shown in Figures 1 (hSBP1) and 2A and 2B (hSBP2). In accordance with the invention, any nucleic acid sequence that encodes an amino acid sequence of an hSBP polypeptide can be used to generate  
5 recombinant molecules which express an hSBP polypeptide. In specific embodiments described herein, a nucleotide sequence encoding a portion of hSBP1 was first isolated as Incyte Clone 606491 from a breast tumor cell line cDNA library BRSTTUT01; and a nucleotide sequence encoding a portion of hSBP2 was first isolated as Incyte Clone 602615 from a breast tumor cell line cDNA library BRSTTUT01.

10 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of degenerate variants of hSBP-encoding nucleotide sequences, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, can be produced. The invention contemplates each and every possible variation of nucleotide sequence that can be made by selecting combinations based on possible codon  
15 choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring hSBP, and all such variations are to be considered as being specifically disclosed herein.

Although nucleotide sequences that encode hSBP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring hSBP under appropriately  
20 selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding hSBP or its derivatives possessing a substantially different codon usage. Codons can be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic expression host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence  
25 encoding hSBP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties (e.g., increased half-life) than transcripts produced from the naturally occurring sequence.

It is now possible to produce a nucleotide sequence encoding an hSBP polypeptide and/or its derivatives entirely by synthetic chemistry, after which the synthetic gene can be inserted into  
30 any of the many available DNA vectors and expression systems using reagents that are well known in the art at the time of the filing of this application. Moreover, synthetic chemistry can be used to introduce mutations into a sequence encoding an hSBP polypeptide.

Also included within the scope of the present invention are polynucleotide sequences that are capable of hybridizing to the nucleotide sequences of Figures 1A-B and/or 2A-B under various conditions of stringency. Hybridization conditions are based on the melting temperature ( $T_m$ ) of the nucleic acid binding complex or probe, as taught in Berger and Kimmel (1987, 5 Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA) incorporated herein by reference, and can be used at a defined stringency.

Altered nucleic acid sequences encoding hSBP that can be used in accordance with the invention include deletions, insertions or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent hSBP. The protein can also 10 comprise deletions, insertions or substitutions of amino acid residues that result in a polypeptide that is functionally equivalent to hSBP. Deliberate amino acid substitutions can be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues with the proviso that biological activity of hSBP is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively 15 charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine; glycine, alanine; asparagine, glutamine; serine, threonine, phenylalanine, and tyrosine.

Alleles of hSBP are also encompassed by the present invention. As used herein, an "allele" or "allelic sequence" is an alternative form of hSBP. Alleles result from a mutation (i.e., 20 an alteration in the nucleic acid sequence) and generally produce altered mRNAs and/or polypeptides that may or may not have an altered structure or function relative to naturally-occurring hSBP. Any given gene may have none, one, or many allelic forms. Common mutational changes that give rise to alleles are generally ascribed to natural deletions, additions or substitutions of amino acids. Each of these types of changes may occur alone or in 25 combination with the other changes, and may occur once or multiple times in a given sequence.

Methods for DNA sequencing are well known in the art and employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland OH), Taq polymerase (Perkin Elmer, Norwalk CT), thermostable T7 polymerase (Amersham, Chicago IL), or combinations of recombinant polymerases and proofreading exonucleases such as the 30 ELONGASE Amplification System marketed by Gibco BRL (Gaithersburg MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown MA) and the ABI 377 DNA

sequencers (Perkin Elmer).

### Extending the Polynucleotide Sequence

The polynucleotide sequence encoding hSBP can be extended utilizing partial nucleotide sequence and various methods known in the art to detect upstream sequences such as promoters and regulatory elements. Clones that contain extended sequences are designated by a suffix (see the tables above). Gobinda et al (1993; PCR Methods Applic 2:318-22) disclose "restriction-site" polymerase chain reaction (PCR) as a direct method which uses universal primers to retrieve unknown sequence adjacent to a known locus. First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR can be used to amplify or extend sequences using divergent primers based on a known region (Triglia T et al (1988) Nucleic Acids Res 16:8186). The primers can be designed using OLIGO® 4.06 Primer Analysis Software (1992; National Biosciences Inc, Plymouth MN), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. This method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Capture PCR (Lagerstrom M et al (1991) PCR Methods Applic 1:111-19) is a method for PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. Capture PCR also requires multiple restriction enzyme digestions and ligations to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before PCR.

Another method that can be used to retrieve unknown sequences is that of Parker JD et al (1991; Nucleic Acids Res 19:3055-60). Additionally, one can use PCR, nested primers, and PromoterFinder libraries to "walk in" genomic DNA (PromoterFinder™ Clontech (Palo Alto CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions. Preferably, the libraries used to identify full length cDNAs have been size-selected to include larger cDNAs. More preferably, the cDNA libraries used to identify full-length cDNAs are those generated using random primers, in that such libraries will contain more sequences

comprising regions 5' of the sequence(s) of interest. A randomly primed library can be particularly useful where oligo d(T) libraries do not yield a full-length cDNA. Genomic libraries are preferred for identification and isolation of 5' nontranslated regulatory regions of a sequence(s) of interest.

- 5           Capillary electrophoresis can be used to analyze the size of, or confirm the nucleotide sequence of, sequencing or PCR products. Systems for rapid sequencing are available from Perkin Elmer, Beckman Instruments (Fullerton CA), and other companies. Capillary sequencing can employ flowable polymers for electrophoretic separation, four different, laser-activatable fluorescent dyes (one for each nucleotide), and a charge coupled device camera for detection of
- 10   the wavelengths emitted by the fluorescent dyes. Output/light intensity is converted to electrical signal using appropriate software (e.g. Genotyper™ and Sequence Navigator™ from Perkin Elmer). The entire process from loading of the samples to computer analysis and electronic data display is computer controlled. Capillary electrophoresis is particularly suited to the sequencing of small pieces of DNA that might be present in limited amounts in a particular sample.
- 15   Capillary electrophoresis provides reproducible sequencing of up to 350 bp of M13 phage DNA in 30 min (Ruiz-Martinez MC et al (1993) Anal Chem 65:2851-2858).

#### **Expression of the Nucleotide Sequence**

- In accordance with the present invention, polynucleotide sequences that encode hSBP polypeptides (which polypeptides include fragments of the naturally-occurring polypeptide.
- 20   fusion proteins, and functional equivalents thereof) can be used in recombinant DNA molecules that direct the expression of hSBP in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence, can be used to clone and express hSBP. As will be understood by those of skill in the art, it may be advantageous to produce hSBP-encoding nucleotide
- 25   sequences possessing non-naturally occurring codons. Codons preferred by a particular prokaryotic or eukaryotic host (Murray E et al (1989) Nuc Acids Res 17:477-508) can be selected, for example, to increase the rate of hSBP expression or to produce recombinant RNA transcripts having a desirable characteristic(s) (e.g., longer half-life than transcripts produced from naturally occurring sequence).

- 30           The nucleotide sequences of the present invention can be engineered in order to alter an hSBP coding sequence for a variety of reasons, including but not limited to, alterations that facilitate the cloning, processing and/or expression of the gene product. For example, mutations

can be introduced using techniques that are well known in the art. e.g., site-directed mutagenesis to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, etc.

In another embodiment of the invention, a natural, modified, or recombinant polynucleotide encoding an hSBP polypeptide can be ligated to a heterologous sequence to encode a fusion protein. For example, where an hSBP polypeptide is to be used in a peptide library for screening and identification of inhibitors of hSBP activity, it may be desirable to provide the hSBP polypeptide in the peptide library as a chimeric hSBP protein that can be recognized by a commercially available antibody. A fusion protein can also be engineered to contain a cleavage site located between an hSBP polypeptide-encoding sequence and a heterologous polypeptide sequence, such that the hSBP polypeptide can be cleaved and purified away from the heterologous moiety.

In an alternative embodiment of the invention, a nucleotide sequence encoding an hSBP polypeptide can be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers et al (1980) Nuc Acids Res Symp Ser 215-23, Horn et al(1980) Nuc Acids Res Symp Ser 225-32, etc). Alternatively, the polypeptide itself can be produced using chemical methods to synthesize an hSBP amino acid sequence, in whole or in part. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge et al (1995) Science 269:202-204) and automated synthesis can be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer) in accordance with the instructions provided by the manufacturer.

The newly synthesized peptide can be substantially purified by preparative high performance liquid chromatography (e.g., Creighton (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co, New York NY). The composition of the synthetic peptides can be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally the amino acid sequence of hSBP, or any part thereof, can be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

#### Expression Systems

In order to express a biologically active hSBP polypeptide, the nucleotide sequence encoding an hSBP polypeptide or its functional equivalent, is inserted into an appropriate expression vector, i.e., a vector having the necessary elements for the transcription and translation



of the inserted coding sequence.

Methods well known to those skilled in the art can be used to construct expression vectors comprising an hSBP polypeptide-encoding sequence and appropriate transcriptional or translational controls. These methods include in vitro recombinant DNA techniques, synthetic techniques and in vivo recombination or genetic recombination. Such techniques are described in Sambrook et al (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY and Ausubel FM et al (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY.

A variety of expression vector/host systems can be utilized to express an hSBP polypeptide-encoding sequence. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transfected with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with bacterial expression vectors (e.g., Ti or pBR322 plasmid); or animal cell systems.

The "control elements" or "regulatory sequences" of these systems, which vary in their strength and specificities, are those nontranslated regions of the vector, enhancers, promoters, and 3' untranslated regions that interact with host cellular proteins to facilitate transcription and translation of a nucleotide sequence of interest. Depending on the vector system and host utilized, any number of suitable transcriptional and translational elements, including constitutive and inducible promoters, can be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, La Jolla CA) or pSport1 (Gibco BRL), ptrp-lac hybrids, and the like can be used. The baculovirus polyhedron promoter can be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) can be cloned into the vector. In mammalian cell systems, promoters from the mammalian genes or from mammalian viruses are most appropriate. Where it is desirable to generate a cell line containing multiple copies of an hSBP polypeptide-encoding sequence, vectors derived from SV40 or EBV can be used in conjunction with other optional vector elements, e.g., an appropriate selectable marker.

In bacterial systems, a number of expression vectors can be used to express an hSBP polypeptide of interest, and will vary with a variety of factors including the intended use intended

for the hSBP polypeptide produced. For example, when large quantities of an hSBP polypeptide are required (e.g., for the antibody production), vectors that direct high-level expression of fusion proteins that can be readily purified may be desirable. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as Bluescript® (Stratagene: 5 which provides for in-frame ligation of a hSBP polypeptide-encoding sequence with sequences encoding the amino-terminal Met and the subsequent 7 residues of  $\beta$ -galactosidase, thereby producing an hSBP polypeptide- $\beta$ -galactosidase hybrid protein); pIN vectors (Van Heeke & Schuster (1989) J Biol Chem 264:5503-5509); and the like. pGEX vectors (Promega, Madison WI) can also be used to express foreign polypeptides as glutathione S-transferase (GST) fusion 10 proteins. In general, such GST fusion proteins are soluble and can be easily purified from cell lysates by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. GST fusion proteins can be designed to include heparin, thrombin or factor XA protease cleavage sites so that the cloned polypeptide of interest can be readily separated from the GST moiety.

15 Where the host cell is yeast (e.g., *Saccharomyces cerevisiae*) a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH can be used. For reviews, see Ausubel et al (supra) and Grant et al (1987) Methods in Enzymology 153:516-544.

Where plant expression vectors are used, the expression of an hSBP polypeptide-encoding 20 sequence can be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV (Brisson et al (1984) Nature 310:511-514) can be used alone or in combination with the omega leader sequence from TMV (Takamatsu et al (1987) EMBO J 6:307-311). Alternatively, plant promoters, such as the small subunit of RUBISCO (Coruzzi et al (1984) EMBO J 3:1671-1680; Broglie et al (1984) Science 224:838-843) or heat 25 shock promoters (Winter J and Sinibaldi RM (1991) Results Probl Cell Differ 17:85-105), can be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. For reviews of such techniques, see Hobbs S or Murry LE in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill New York NY, pp 191-196 or Weissbach and Weissbach (1988) Methods for Plant Molecular Biology, Academic 30 Press, New York NY, pp 421-463.

Alternatively, insect cell expression systems can be used to express an hSBP polypeptide.. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a

vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The hSBP polypeptide-encoding sequence can be cloned into a nonessential region of the virus, such as the polyhedron gene, and placed under control of the polyhedron promoter. Successful insertion of hSBP renders the polyhedron gene inactive and produces recombinant virus lacking coat protein. The recombinant viruses are then used to infect S. frugiperda cells or Trichoplusia larvae for expression of hSBP polypeptide (Smith et al (1983) J Virol 46:584; Engelhard EK et al (1994) Proc Nat Acad Sci 91:3224-7).

Where the host cell is a mammalian cells, a number of viral-based expression systems can be used. For example, the expression vector can be derived from an adenovirus nucleotide sequence. An hSBP polypeptide-encoding sequence can be ligated into an adenovirus transcription/translation complex, which is composed of the late promoter and tripartite leader sequence. Insertion of the nucleotide sequence of interest into a nonessential E1 or E3 region of the viral genome will result in the production of a viable virus capable of expressing hSBP polypeptide in infected host cells (Logan and Shenk (1984) Proc Natl Acad Sci 81:3655-59). In addition, transcriptional enhancers, such as the Rous sarcoma virus (RSV) enhancer, can be used to increase expression in mammalian host cells.

Specific initiation signals may also be required for efficient translation of an hSBP polypeptide-encoding sequence, e.g., the ATG initiation codon and flanking sequences. Where a native hSBP polypeptide encoding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, where only coding sequence, or a portion thereof, is inserted in an expression vector, exogenous transcriptional control signals including the ATG initiation codon must be provided. Furthermore, the initiation codon must be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be derived from various origins, and can be either natural or synthetic. Expression efficiency can be enhanced by including enhancers appropriate to the cell system in use (Scharf D et al (1994) Results Probl Cell Differ 20:125-62; Bittner et al (1987) Methods in Enzymol 153:516-544).

Host cells can be selected for hSBP polypeptide expression according to the ability of the cell to modulate the expression of the inserted sequences or to process the expressed protein in a desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing that involves cleavage of a "prepro" form of the protein may also be

important for correct polypeptide folding, membrane insertion, and/or function. Host cells such as CHO, HeLa, MDCK, 293, WI38, and others have specific cellular machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing of the introduced, foreign polypeptide.

5           Where long-term, high-yield recombinant polypeptide production is desired, stable expression is preferred. For example, cell lines that stably express hSBP can be transformed using expression vectors containing viral origins of replication or endogenous expression elements and a selectable marker gene. After introduction of the vector, cells can be grown for 1-2 days in an enriched media before they are exposed to selective media. The selectable marker, 10 which confers resistance to the selective media, allows growth and recovery of cells that successfully express the introduced sequences. Resistant, stably transformed cells can be proliferated using tissue culture techniques appropriate to the host cell type.

Any number of selection systems can be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler M et al (1977) 15 Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy I et al (1980) Cell 22:817-23) genes which can be employed in tk- or aprt- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler M et al (1980) Proc Natl Acad Sci 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin F et al (1981) J 20 Mol Biol 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman SC and RC Mulligan (1988) Proc Natl Acad Sci 85:8047-51). Recently, the use of visible markers has gained popularity with such 25 markers as anthocyanins,  $\beta$ -glucuronidase and its substrate, GUS, and luciferase and its substrate, luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes CA et al (1995) Methods Mol Biol 55:121-131).

#### **Identification of Transformants Containing the Polynucleotide Sequence**

30           Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression should be confirmed. For example, if the hSBP polypeptide encoding sequence is inserted within a marker gene sequence, recombinant

cells containing this sequence can be identified by the absence of marker gene function.

Alternatively, a marker gene can be placed in tandem with a hSBP sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection is indicative of expression of the tandem hSBP.

5 Alternatively, host cells that contain the coding sequence for hSBP polypeptides and express hSBP polypeptides can be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridization and protein bioassay or immunoassay techniques including membrane, solution, or chip-based technologies for the detection and/or quantitation of the nucleic acid or protein.

10 The presence of the polynucleotide sequence encoding hSBP polypeptides can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes, portions or fragments of polynucleotides encoding hSBP. Nucleic acid amplification-based assays involve the use of oligonucleotides or oligomers based on the hSBP polypeptide-encoding sequence to detect transformants containing hSBP polypeptide-encoding DNA or RNA. As used herein

15 "oligonucleotides" or "oligomers" refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides which can be used as a probe or amplimer.

A variety of protocols for detecting and measuring the expression of hSBP, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples  
20 include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on hSBP is preferred, but a competitive binding assay can be employed. These and other assays are described in, e.g., Hampton R et al (1990, Serological Methods, a Laboratory Manual, APS Press, St Paul MN) and Maddox DE et al  
25 (1983, J Exp Med 158:1211).

A wide variety of detectable labels and conjugation techniques are known by in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to hSBP-encoding polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled  
30 nucleotide. Alternatively, an nucleotide sequence encoding an hSBP polypeptide can be cloned into a vector for the production of an mRNA probe. Such vectors, which are known in the art and commercially available, can be used to synthesize RNA probes in vitro by addition of an

appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides.

A number of companies, including Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH), supply commercial kits and protocols suitable for the methods described above. Suitable reporter molecules or labels include those  
5 radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like, as described in U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241, each of which are incorporated herein by reference. Recombinant immunoglobulins can be produced as according to U.S. Patent No. 4,816,567, incorporated herein by reference.

#### 10 **Purification of hSBP**

Host cells transformed with a nucleotide sequence encoding an hSBP polypeptide can be cultured under conditions suitable for the expression and recovery of the hSBP polypeptide from cell culture. The polypeptide produced by a recombinant cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those  
15 of skill in the art, expression vectors containing polynucleotides encoding hSBP polypeptides can be designed with signal sequences that direct secretion of hSBP through a prokaryotic or eukaryotic cell membrane.

Recombinant hSBP constructs can also include a nucleotide sequence(s) encoding one or more polypeptide domains that, when expressed in-frame with the hSBP-encoding sequence,  
20 facilitates purification of soluble proteins (Kroll DJ et al (1993) DNA Cell Biol 12:441-53; c.f. discussion of vectors infra containing fusion proteins). Such purification facilitating domains include, but are not limited to, metal chelating peptides (e.g., histidine-tryptophan modules) that allow purification with immobilized metals, protein A domains that allow purification with immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity  
25 purification system (Immunex Corp, Seattle WA). A cleavable linker sequences(s) (e.g., Factor XA or enterokinase (Invitrogen, San Diego CA)) between the purification domain and the hSBP polypeptide-encoding sequence can be included to facilitate purification. One such expression vector provides for expression of a fusion protein comprising 6 histidine residues followed by thioredoxin and an enterokinase cleavage site. The histidine residues facilitate purification on  
30 IMIAC (immobilized metal ion affinity chromatography as described in Porath et al (1992) Protein Expression and Purification 3: 263-281), while the enterokinase cleavage site provides a means for separating the hSBP domain from the remainder of the fusion protein.

hSBP polypeptides (which polypeptides encompass polypeptides composed of a portion of the native hSBP amino acid sequence) can also be produced by direct peptide synthesis using solid-phase techniques (cf Stewart et al (1969) Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco; Merrifield J (1963) J Am Chem Soc 85:2149-2154). *In vitro* protein synthesis  
5 can be performed using manual techniques or by automation. Automated synthesis can be achieved by, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City CA) in accordance with the instructions provided by the manufacturer. Various fragments of hSBP can be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

#### 10 Uses of hSBP

The rationale for use of the nucleotide and polypeptide sequences disclosed herein is based in part on the differential expression of hSBP-encoding sequences in breast tumor tissue and in part on the chemical and structural homology between the hSBP proteins disclosed herein and chemical and structural homology between: 1) hSBP1, rat prostatic binding proteins C1  
15 (GI 206442; Delaey et al. supra), rat prostatic binding protein C2(Delaey et al. 1987 Nucl Acid Res 15:1627-1641) and rabbit uteroglobin (Menne et al. 1982 Proc Natl Acad Sci USA 79:4853-4857) (Figure 5), and 2) hSBP2, human mammaglobin (GI 1199595; Watson et al. supra), and rat prostatic binding protein C3 (GI 206543; Parker et al. supra) (Figure 6).

Accordingly, hSBP or an hSBP derivative can be used in the diagnosis and management  
20 of breast cancer. Given the homology of hSBP with rat PBP, and the differential expression of hSBP in human breast tumor tissue, hSBP can be used as a diagnostic marker for human breast cancer. Expression of rat PBP is regulated by androgens (Muder et al. 1984 Biochem Biophys Acta 781:121-9; Page et al. 1983 Cell 32:495-502) and by growth hormone (Reiter et al. 1995 Endocrinol 166: 3338-44). Thus the level of hSBP can serve as a marker for transformation of  
25 normal breast cells into cancerous cells. Alternatively, or in addition, development of breast cancer can be detected by examining the ratio of hSBP to the levels of steroid hormones (e.g., testosterone or estrogen) or to other hormones (e.g., growth hormone, insulin). Thus expression of hSBP1 and/or hSBP2 can also be used to discriminate between normal and cancerous breast tissue, to discriminate between different types of breast cancer, to provide guidance in selection  
30 of anti-cancer therapies, to monitor the progress of patients undergoing chemotherapy and/or other anti-cancer treatments, to determine the success of surgery to remove cancerous tissue, and to monitor patients who have had or are susceptible to breast cancer. In addition to diagnosis and

treatment of breast cancer after its development, detection of hSBP expression can be used to identify patients susceptible to breast cancer. Expression of hSBP in cancerous cells can be examined in breast tissue in situ or in pathology sections. Alternatively, if hSBP is secreted at sufficient levels, expression of hSBP can be assessed in blood, serum, or plasma. Assessment of levels of hSBP expression can be used to differentiate between normal and cancerous breast tissue, and/or different types of cancerous breast tissue (e.g., invasive vs. non-invasive; ductal vs. axillary lymph node). In addition, because hSBP is differentially expressed in breast tumor cells, hSBP polypeptides can serve as a target for anti-cancer therapy that is targeted to hSBP-expressing breast tumor cells. For example, cells can be transfected with antisense sequences to hSBP-encoding polynucleotides or provided with antagonists to hSBP to reduce or eliminate hSBP expression in cancerous breast cells. Alternatively, cancerous breast cells, or breast cells susceptible to cancer, can be transformed (e.g., via gene therapy techniques) with hSBP-encoding nucleic acid to provide for expression of excess hSBP and interruption of steroid binding.

#### **hSBP Antibodies**

hSBP-specific antibodies are useful for the diagnosis of conditions and diseases associated with expression of hSBP. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library. Neutralizing antibodies, i.e., those which inhibit a biochemical activity of hSBP, are especially preferred for diagnostics and therapeutics.

hSBP polypeptides suitable for production of antibodies need not be biologically active; rather, the polypeptide, or oligopeptide need only be antigenic. Polypeptides used to generate hSBP-specific antibodies generally have an amino acid sequence consisting of at least five amino acids, preferably at least 10 amino acids. Preferably, antigenic hSBP polypeptides mimic an epitope of the native hSBP. Antibodies specific for short hSBP polypeptides can be generated by linking the hSBP polypeptide to a carrier, or fusing the hSBP polypeptide to another protein (e.g., keyhole limpet hemocyanin), and using the carrier-linked or hSBP chimeric molecule as an antigen. In general, anti-hSBP antibodies can be produced according to methods well known in the art.

Various hosts, generally mammalian hosts, can be used to produce anti-hSBP antibodies (e.g., goats, rabbits, rats, mice). Anti-hSBP antibodies are produced by immunizing the host (e.g., by injection) with an hSBP polypeptide that retains immunogenic properties (which encompasses any portion of native hSBP, fragment or oligopeptide). Depending on the host



species, various adjuvants can be used to increase the host's immunological response. Such adjuvants include but are not limited to, Freund's, mineral gels (e.g., aluminum hydroxide), and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are potentially useful human adjuvants.

Monoclonal anti-hSBP antibodies can be prepared using any technique that provides for the production of antibody molecules by immortalized cell lines in culture. These techniques include, but are not limited to, the hybridoma technique originally described by Koehler and Milstein (1975 Nature 256:495-497), the human B-cell hybridoma technique (Kosbor et al (1983) Immunol Today 4:72; Cote et al (1983) Proc Natl Acad Sci 80:2026-2030) and the EBV-hybridoma technique (Cole et al (1985) Monoclonal Antibodies and Cancer Therapy, Alan R Liss Inc, New York NY, pp 77-96).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison et al (1984) Proc Natl Acad Sci 81:6851-6855; Neuberger et al (1984) Nature 312:604-608; Takeda et al (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce hSBP-specific single chain antibodies

Antibodies can be produced in vivo or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi et al (1989, Proc Natl Acad Sci 86: 3833-3837), and Winter G and Milstein C (1991; Nature 349:293-299).

Antibody fragments having specific binding sites for an hSBP polypeptide can also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments, which can be produced by pepsin digestion of the antibody molecule, and Fab fragments, which can be generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries can be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse WD et al (1989) Science 256:1275-1281).

A variety of protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies having established antigen specificities are well known in the art. Such immunoassays typically involve the formation of complexes between an hSBP polypeptide and a specific anti-hSBP antibody, and the detection and quantitation of hSBP-antibody complex formation. A two-site, monoclonal-based immunoassay utilizing monoclonal

antibodies reactive to two noninterfering epitopes on a specific hSBP protein is preferred, but a competitive binding assay can also be employed. These assays are described in Maddox DE et al (1983, J Exp Med 158:1211).

#### Diagnostic Assays Using hSBP Specific Antibodies

5           Particular hSBP antibodies are useful for the diagnosis of conditions or diseases characterized by expression of hSBP (e.g., breast cancer) or in assays to monitor patients being treated with hSBP, agonists, antagonists, or inhibitors. Diagnostic assays for hSBP include methods using a detectably-labeled anti-hSBP antibody to detect hSBP in human body fluids or extracts of cells or tissues. The polypeptides and antibodies of the present invention can be used  
10           with or without modification. Frequently, the polypeptides and antibodies are labeled by covalent or noncovalent attachment to a reporter molecule. A wide variety of such suitable reporter molecules are known in the art.

          A variety of protocols for detection and quantifying hSBP, using either polyclonal or monoclonal antibodies specific for an hSBP polypeptide, are known in the art. Examples include  
15           enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on hSBP is preferred, but a competitive binding assay can instead be employed. These assays are described, among other places, in Maddox, DE et al (1983, J Exp Med 158:1211).

20           In order to provide a basis for diagnosis, normal or standard values for hSBP expression must be established. This is accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, preferably human, with antibody to hSBP under conditions suitable for complex formation according to methods well known in the art. The amount of standard complex formation can be quantified by comparing detection levels  
25           associated with known quantities of hSBP with detection levels associated with both control and disease samples from biopsied tissues. Standard values obtained from normal samples are compared with values obtained from samples from subjects potentially affected by disease. Deviation between standard and subject values establishes the presence of disease state.

#### Drug Screening

30           hSBP polypeptides, which encompass biologically active or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any of a variety of drug screening techniques. The polypeptide employed in such a test can be free in solution,

affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between hSBP and the agent being tested, can be measured.

Preferably, the drug screening technique used provides for high throughput screening of compounds having suitable binding affinity to the hSBP, as described in detail in "Determination of Amino Acid Sequence Antigenicity" by Geysen HN. WO Application 84/03564, published on September 13, 1984, and incorporated herein by reference. In summary, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with hSBP polypeptides, unreacted materials are washed away, and bound hSBP is detected by methods well known in the art.

Purified hSBP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the polypeptide and immobilize it on a solid support.

The invention also contemplates the use of competitive drug screening assays in which hSBP-specific neutralizing antibodies compete with a test compound for binding of hSBP polypeptide. In this manner, the antibodies can be used to detect the presence of any polypeptide that shares one or more antigenic determinants with an hSBP polypeptide.

#### **Uses of the Polynucleotide Encoding hSBP**

A polynucleotide encoding an hSBP polypeptide (which polypeptides include native hSBP and fragments thereof) can be used for diagnostic and/or therapeutic purposes. For diagnostic purposes, polynucleotides encoding hSBP of this invention can be used to detect and quantitate gene expression in biopsied tissues in which expression of hSBP is implicated, particularly in diagnosis of breast cancer. The diagnostic assay is useful to assess hSBP expression levels (e.g., to distinguish between the absence, and presence of hSBP expression, as well as to assess various hSBP expression levels (e.g., excessively high, high, moderate, or low)) and to monitor regulation of hSBP levels during therapeutic intervention. Included in the scope of the invention are oligonucleotide sequences, antisense RNA and DNA molecules, and peptide nucleic acids (PNAs).

Another aspect of the subject invention is to provide for hybridization or PCR probes capable of detecting polynucleotide sequences encoding hSBP, including genomic sequences and closely related molecules. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' region, and the stringency of the hybridization or amplification (maximal,

high, intermediate or low) will determine whether the probe identifies only naturally occurring sequences encoding hSBP, alleles or related sequences.

The probes of the invention can be used in the detection of related sequences; such probes preferably comprise at least 50% of the nucleotides from any of the hSBP polypeptide-encoding sequences described herein. The hybridization probes of the subject invention can be derived from the nucleotide sequence of SEQ ID NO:2 and SEQ ID NO:4, or from their corresponding genomic sequences including promoters, enhancer elements and introns of the naturally occurring hSBP-encoding sequences. Hybridization probes can be detectably labeled with a variety of reporter molecules, including radionuclides (e.g., <sup>32</sup>P or <sup>35</sup>S), or enzymatic labels (e.g., alkaline phosphatase coupled to the probe via avidin/biotin coupling systems), and the like.

Specific hybridization probes for hSBP-encoding DNAs can also be produced by cloning nucleic acid sequences encoding hSBP or hSBP derivatives into vectors for production of mRNA probes. Such vectors, which are known in the art and are commercially available, can be used to synthesize RNA probes *in vitro* using an appropriate RNA polymerase (e.g., T7 or SP6 RNA polymerase) and appropriate radioactively labeled nucleotides.

#### **Diagnostic Use**

Polynucleotide sequences encoding hSBP polypeptide can be used in the diagnosis of conditions or diseases associated with hSBP expression, especially breast cancer. For example, polynucleotide sequences encoding hSBP can be used in hybridization or PCR assays of fluids or tissues from biopsies to detect hSBP expression. Suitable qualitative or quantitative methods include Southern or northern analysis, dot blot or other membrane-based technologies; PCR technologies; dip stick, pIN, chip and ELISA technologies. All of these techniques are well known in the art and are the basis of many commercially available diagnostic kits.

The nucleotide sequences encoding hSBP disclosed herein provide the basis for assays that detect the onset of, susceptibility to, or the presence of breast cancer. Nucleotide sequences encoding hSBP polypeptides can be labeled by methods known in the art and combined with a fluid or tissue sample from a patient suspected of having or susceptible to breast cancer under conditions suitable for the formation of hybridization complexes. After an incubation period, the sample is washed with a compatible fluid which optionally contains a dye (or other label requiring a developer) if the nucleotide has been labeled with an enzyme. After the compatible fluid is rinsed off, the dye is quantitated and compared with a standard. If the amount of dye in the biopsied or extracted sample is significantly elevated over that of a comparable negative

control sample, the nucleotide sequence has hybridized with nucleotide sequences in the sample. The presence of hSBP-encoding nucleotide sequences in the sample, particularly the presence of elevated levels of hSBP-encoding sequences, indicates that the patient has or is at risk of developing the associated disease.

5        Such assays can also be used to evaluate the efficacy of a particular therapeutic treatment regime in animal studies or in clinical trials, or in monitoring the treatment of an individual patient. In order to provide a basis for the diagnosis of disease, a normal or standard profile for hSBP expression must be established. This is accomplished by combining body fluids or cell  
10        extracts taken from normal subjects, either animal or human, with hSBP, or a portion thereof, under conditions suitable for hybridization or amplification. Standard hybridization can be quantified by comparing, in the same experiment, the values obtained for normal subjects with those obtained with a dilution series of hSBP containing known amounts of substantially purified hSBP. Standard values obtained from normal samples are compared with values  
15        obtained from samples from patients afflicted with hSBP-associated diseases, or suspected of having such diseases (e.g., breast cancer). Deviation between standard and subject values is used to establish the presence of disease.

Once disease is established, a therapeutic agent is administered and a treatment profile is generated. Such assays can be repeated on a regular basis to evaluate whether the values in the profile progress toward or return to a normal or standard pattern of hSBP expression. Successive  
20        treatment profiles can be used to show the efficacy of treatment over a period of several days or several months.

Oligonucleotides based upon hSBP sequences can be used in PCR-based techniques, as described in U.S. Patent Nos. 4,683,195 and 4,965,188. Such oligomers are generally chemically synthesized, or produced enzymatically or by recombinantly. Oligomers generally comprise two  
25        nucleotide sequences, one with sense orientation (5'→3') and one with antisense (3'←5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers can be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

30        Additional methods for quantitation of expression of a particular molecule according to the invention include radiolabeling (Melby PC et al 1993 J Immunol Methods 159:235-44) or biotinylating (Duplaa C et al 1993 Anal Biochem 229:36) nucleotides, coamplification of a

control nucleic acid, and interpolation of experimental results according to standard curves. Quantitation of multiple samples can be made more time efficient by running the assay in an ELISA format in which the oligomer of interest is presented in various dilutions and rapid quantitation is accomplished by spectrophotometric or colorimetric detection. For example, the presence of a relatively high amount of hSBP in extracts of biopsied tissues indicates the presence of cancerous breast cells. A definitive diagnosis of this type can allow health professionals to begin aggressive treatment and prevent further worsening of the condition. Similarly, further assays can be used to monitor the progress of a patient during treatment. Furthermore, the nucleotide sequences disclosed herein can be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known such as the triplet genetic code, specific base pair interactions, and the like:

#### Therapeutic Use

Based upon its homology to genes encoding prostatic binding proteins, hSBP polypeptides and its expression profile in breast tumor cells, polynucleotide sequences encoding hSBP disclosed herein may be useful in the treatment of conditions such as breast cancer or other condition associated with hSBP expression or over-expression.

Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids, can be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Recombinant vectors for expression of antisense hSBP polynucleotides can be constructed according to methods well known in the art (see, for example, the techniques described in Sambrook et al (supra) and Ausubel et al (supra)).

Polynucleotides comprising the full length cDNA sequence and/or its regulatory elements enable researchers to use sequences encoding hSBP as an investigative tool in sense (Yousoufian H and HF Lodish 1993 Mol Cell Biol 13:98-104) or antisense (Eguchi et al (1991) Ann Rev Biochem 60:631-652) regulation of gene function. Such technology is now well known in the art, and sense or antisense oligomers, or larger fragments, can be designed from various locations along the coding or control regions.

Expression of genes encoding hSBP can be decreased by transfecting a cell or tissue with expression vectors that express high levels of a desired hSBP-encoding fragment. Such constructs can flood cells with untranslatable sense or antisense sequences. Even in the absence of integration into the DNA, such vectors can continue to transcribe RNA molecules until all

copies are disabled by endogenous nucleases. Transient expression can last for a month or more with a non-replicating vector (Mettler I, personal communication) and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA or PNA, to the control regions of gene encoding hSBP (i.e., the promoters, enhancers, and introns). Oligonucleotides derived from the transcription initiation site, e.g., between -10 and +10 regions of the leader sequence, are preferred. The antisense molecules can also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Similarly, inhibition of expression can be achieved using "triple helix" base-pairing methodology. Triple helix pairing compromises the ability of the double helix to open sufficiently for binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA were reviewed by Gee JE et al (In: Huber BE and BI Carr (1994) Molecular and Immunologic Approaches, Futura Publishing Co, Mt Kisco NY).

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. The invention contemplates engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding hSBP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, which sites include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences between 15 and 20 ribonucleotides corresponding to a region of the target gene containing the cleavage site can be evaluated for secondary structural features that can render the oligonucleotide inoperable. The suitability of candidate targets can also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Antisense molecules and ribozymes of the invention can be prepared by methods known in the art for the synthesis of RNA molecules, including techniques for chemical oligonucleotide synthesis, e.g., solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules can be generated by in vitro and in vivo transcription of DNA sequences encoding hSBP. Such DNA sequences can be incorporated into a wide variety of vectors with suitable RNA polymerase promoters (e.g, T7 or SP6). Alternatively, antisense cDNA constructs useful in the constitutive or inducible synthesis of antisense RNA can be introduced into cell lines, cells, or tissues.

RNA molecules can be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine and wybutosine as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine that are not as easily recognized by endogenous endonucleases.

Methods for introducing vectors into cells or tissues include those methods discussed *infra* and which are equally suitable for *in vivo*, *in vitro* and *ex vivo* therapy. In *ex vivo* therapy, vectors are introduced into stem cells obtained from the patient and clonally propagated for autologous transplant back into that same patient (see, e.g., U.S. Patent Nos. 5,399,493 and 5,437,994, incorporated herein by reference). Transfection and by liposome methods for delivery of a nucleotide sequence of interest to accomplish gene therapy are well known in the art.

Furthermore, the nucleotide sequences for hSBP disclosed herein can be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including but not limited to such properties as the triplet genetic code and specific base pair interactions.

#### **Detection and Mapping of Related Polynucleotide Sequences**

The hSBP nucleic acid sequences can also be used to generate hybridization probes for mapping the naturally occurring genomic sequence. The sequence can be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. These include *in situ* hybridization to chromosomal spreads, flow-sorted chromosomal preparations, or artificial chromosome constructions such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price CM (1993; Blood Rev 7:127-34) and Trask BJ (1991; Trends Genet 7:149-54).

The technique of fluorescent *in situ* hybridization of chromosome spreads is described in, for example, Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY. Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques can be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a gene encoding hSBP on a physical



chromosomal map and a specific disease (or predisposition to a specific disease) can help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention can be used to detect differences in gene sequences between normal, carrier, or affected individuals.

5        In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers can be used for extending genetic maps. For example an sequence tagged site based map of the human genome was recently published by the Whitehead-MIT Center for Genomic Research (Hudson TJ et al (1995) Science 270:1945-1954). Often the placement of a gene on the chromosome of another mammalian  
10 species such as a mouse (Whitehead Institute/MIT Center for Genome Research, Genetic Map of the Mouse, Database Release 10, April 28, 1995) can reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. Physical mapping provides valuable information to investigators searching for disease genes using positional cloning or other  
15 gene discovery techniques. Once a disease or syndrome, such as ataxia telangiectasia (AT), has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti et al (1988) Nature 336:577-580), other sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequence of the subject invention can also be used to detect differences in the chromosomal location due to  
20 translocation, inversion, etc. among normal, carrier or affected individuals.

#### **Pharmaceutical Compositions**

The present invention relates to pharmaceutical compositions which can comprise nucleotides, proteins, antibodies, agonists, antagonists, or inhibitors, alone or in combination with at least one other agent, such as a stabilizing compound, which can be administered in any  
25 sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. Any of these molecules can be administered to a patient alone or in combination with other agents, drugs or hormones, in pharmaceutical compositions where it is mixed with excipient(s), or with pharmaceutically acceptable carriers. In one embodiment of the present invention, the pharmaceutically acceptable carrier is pharmaceutically inert.

#### **Administration of Pharmaceutical Compositions**

Administration of pharmaceutical compositions is accomplished orally or parenterally. Methods of parenteral delivery include topical, intra-arterial (e.g., directly to the breast tumor),

intramuscular, subcutaneous, intramedullary, intrathecal, intraventricular, intravenous, intraperitoneal, or intranasal administration. In addition to the active ingredients, these pharmaceutical compositions can contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries that facilitate processing of the active compounds into preparations for pharmaceutical use. Further details on techniques for formulation and administration can be found in the latest edition of "Remington's Pharmaceutical Sciences" (Maack Publishing Co, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; and gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores are provided with suitable coatings such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations that can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders such as lactose or starches, lubricants such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations for parenteral administration include aqueous solutions of active compounds. For injection, the pharmaceutical compositions of the invention can be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection  
5 suspensions can contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds can be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension can also contain suitable stabilizers or  
10 agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

#### **Manufacture and Storage**

15 The pharmaceutical compositions of the present invention can be manufactured in any suitable manner known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes.

The pharmaceutical composition can be provided as a salt and can be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic,  
20 etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation can be a lyophilized powder in 1mM-50 mM histidine, 0.1%-2% sucrose, 2%-7% mannitol at a pH range of 4.5 to 5.5 that is combined with buffer prior to use.

After pharmaceutical compositions comprising a compound of the invention formulated  
25 in a acceptable carrier have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of hSBP, such labeling would include amount, frequency and method of administration.

#### **Therapeutically Effective Dose**

Pharmaceutical compositions suitable for use in the present invention include  
30 compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model is also used to achieve a desirable concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of protein or its antibodies, antagonists, or inhibitors that ameliorate the symptoms or condition. Therapeutic efficacy and toxicity of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals. e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and expressed as the ratio LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The actual dosage can vary within this range depending upon, for example, the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage is chosen by the individual physician in view of the patient to be treated. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Additional factors that may be taken into account include the severity of the disease state, e.g., tumor size and location; age, weight and gender of the patient; diet, time and frequency of administration; drug combination(s); reaction sensitivities; and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

It is contemplated, for example, that hSBP or an hSBP derivative can be delivered in a

suitable formulation to block the progression of breast cancer. Similarly, administration of hSBP antagonists may also inhibit the activity or shorten the lifespan of this protein.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

5

## INDUSTRIAL APPLICABILITY

### I. Construction of BRSTTUT01 cDNA Libraries

The BRSTTUT01 cDNA library was constructed from breast tumor removed from a 55 year old female (lot #0005; Mayo Clinic, Rochester MN). The frozen tissue was immediately homogenized and lysed using a Brinkmann Homogenizer Polytron-PT 3000 (Brinkmann Instruments, Inc. Westbury NY) in guanidinium isothiocyanate solution. Lysates were then loaded on a 5.7 M CsCl cushion and ultracentrifuged in a SW28 swinging bucket rotor for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted once with acid phenol at pH 4.0 and once with phenol chloroform at pH 8.0 and precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in DEPC-treated water and DNase treated for 25 min at 37°. The reaction was stopped with an equal volume of acid phenol, and the RNA was isolated using the Qiagen Oligotex kit (QIAGEN Inc, Chatsworth CA) and used to construct the cDNA library. The RNA was handled according to the recommended protocols in the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (catalog #18248-013; Gibco/BRL). cDNAs were fractionated on a Sepharose CL4B column (catalog #275105, Pharmacia), and those cDNAs exceeding 400 bp were ligated into pSport I. The plasmid pSport I was subsequently transformed into DH5a(tm) competent cells (Cat. #18258-012, Gibco/BRL).

### II. Isolation and Sequencing of cDNA Clones From BRSTTUT01

Plasmid DNA was released from the cells and purified using the Miniprep Kit (Catalogue # 77468; Advanced Genetic Technologies Corporation, Gaithersburg MD). This kit consists of a 96 well block with reagents for 960 purifications. The recommended protocol was employed except for the following changes: 1) the 96 wells were each filled with only 1 ml of sterile Terrific Broth (Catalog # 22711, LIFE TECHNOLOGIES(tm), Gaithersburg MD) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) the bacteria were cultured for 24 hours after the wells were inoculated and then lysed with 60 µl of lysis buffer; 3) a centrifugation step employing the Beckman GS-6R @2900 rpm for 5 min was performed before the contents of the block were added to the primary filter plate; and 4) the optional step of adding isopropanol to TRIS buffer was not routinely performed. After the last step in the protocol, samples were

30

transferred to a Beckman 96-well block for storage.

The cDNAs were sequenced by the method of Sanger F and AR Coulson (1975; J Mol Biol 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno NV) in combination with four Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown MA) and Applied Biosystems 377 or 373 DNA Sequencing Systems (Perkin Elmer), and reading frame was determined.

### III. Homology Searching of cDNA Clones and Their Deduced Proteins

Each cDNA was compared to sequences in GenBank using a search algorithm developed by Applied Biosystems and incorporated into the INHERIT™ 670 Sequence Analysis System. In this algorithm, Pattern Specification Language (TRW Inc, Los Angeles CA) was used to determine regions of homology. The three parameters that determine how the sequence comparisons run were window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database was searched for sequences containing regions of homology to the query sequence, and the appropriate sequences were scored with an initial value. Subsequently, these homologous regions were examined using dot matrix homology plots to distinguish regions of homology from chance matches. Smith-Waterman alignments were used to display the results of the homology search.

Peptide and protein sequence homologies were ascertained using the INHERIT- 670 Sequence Analysis System in a way similar to that used in DNA sequence homologies. Pattern Specification Language and parameter windows were used to search protein databases for sequences containing regions of homology which were scored with an initial value. Dot-matrix homology plots were examined to distinguish regions of significant homology from chance matches.

BLAST, which stands for Basic Local Alignment Search Tool (Altschul SF (1993) J Mol Evol 36:290-300; Altschul, SF et al (1990) J Mol Biol 215:403-10), was used to search for local sequence alignments. BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs. BLAST is useful for matches that do not contain gaps. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach identifies HSPs between a query sequence

and a database sequence. evaluates the statistical significance of any matches found. and reports only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output.

#### IV. Northern Analysis

Northern analysis. a laboratory technique used to detect the presence of a gene transcript, and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al. supra).

Analogous computer techniques using BLAST (Altschul SF 1993 and 1990, supra) are used to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ™ database (Incyte, Palo Alto CA). This analysis is much faster than multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores can identify related molecules. The abundance data (Abun) represent the number of transcripts of the gene of interest in the cDNA library. Percent abundance is calculated by dividing the number of transcripts of a gene of interest present in a cDNA library by the total number of transcripts in the cDNA library.

#### V. Extension of hSBP-Encoding Polynucleotides to FullLength or to Recover Regulatory Elements

Full length hSBP-encoding nucleic acid sequences (SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6) are used to design oligonucleotide primers for extending a partial nucleotide sequence to full length and/or for obtaining 5' sequences from genomic libraries. One synthesized primer is used to initiate extension in the antisense direction (XLR), and a second synthesized primer is

used to extend sequence in the sense direction (XLF). Primers allow the extension of the known hSBP-encoding sequence "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest (U.S. Patent Application 08/487,112, filed June 7, 1995, specifically incorporated by reference). The initial primers are designed from the cDNA using  
 5 OLIGO® 4.06 Primer Analysis Software (National Biosciences), or another appropriate program. The initial primers are preferable designed to be 22-30 nucleotides in length, have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. Any stretch of nucleotides that would result in hairpin structures and primer-primer dimerizations is avoided.

10 The original, selected cDNA libraries, or a human genomic library, are used to extend the sequence; the latter is most useful to obtain 5' upstream regions. If more extension is necessary or desired, additional sets of primers are designed to further extend the known region.

By following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix, high fidelity amplification is obtained. Beginning with 40 pmol of  
 15 each primer and the recommended concentrations of all other components of the kit, PCR is performed using the Peltier Thermal Cycler (PTC200; MJ Research, Watertown MA) and the following parameters:

	Step 1	94° C for 1 min (initial denaturation)
	Step 2	65° C for 1 min
20	Step 3	68° C for 6 min
	Step 4	94° C for 15 sec
	Step 5	65° C for 1 min
	Step 6	68° C for 7 min
	Step 7	Repeat step 4-6 for 15 additional cycles
25	Step 8	94° C for 15 sec
	Step 9	65° C for 1 min
	Step 10	68° C for 7:15 min
	Step 11	Repeat step 8-10 for 12 cycles
	Step 12	72° C for 8 min
30	Step 13	4° C (and holding)

A 5-10  $\mu$ l aliquot of the reaction mixture is analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands containing the largest products were selected and cut out of the  
 35 gel. Further purification is accomplished using a commercial gel extraction method such as QIAQuick™ (QIAGEN Inc). After recovery of the DNA, Klenow enzyme was used to trim single-stranded, nucleotide overhangs creating blunt ends to facilitate religation and cloning.



After ethanol precipitation, the products are redissolved in 13  $\mu$ l of ligation buffer. 1  $\mu$ l T4-DNA ligase (15 units) and 1  $\mu$ l T4 polynucleotide kinase are added, and the mixture is incubated at room temperature for 2-3 hours or overnight at 16° C. Competent *E. coli* cells (in 40  $\mu$ l of appropriate media) are transformed with 3  $\mu$ l of ligation mixture and cultured in 80  $\mu$ l of SOC medium (Sambrook J et al. supra). After incubation for one hour at 37° C, the whole transformation mixture is plated on Luria Bertani (LB)-agar (Sambrook J et al, supra) containing 2xCarb. The following day, several colonies are randomly picked from each plate and cultured in 150  $\mu$ l of liquid LB/2xCarb medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5  $\mu$ l of each overnight culture is transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5  $\mu$ l of each sample was transferred into a PCR array.

For PCR amplification, 18  $\mu$ l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the

following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions are run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid and sequenced.

#### VI. Labeling and Use of Hybridization Probes

Hybridization probes derived from SEQ ID NO:2 and SEQ ID NO:4 are used to screen cDNAs, genomic DNAs or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base-pairs, is specifically described, essentially the same procedure is used with larger cDNA fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences), labeled by combining 50 pmol of each oligomer and 250 mCi of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham, Chicago IL) and T4 polynucleotide kinase (DuPont NEN®, Boston MA). The labeled oligonucleotides are substantially purified with Sephadex G-25

super fine resin column (Pharmacia). A portion containing  $10^7$  counts per minute of each of the sense and antisense oligonucleotides is used in a typical membrane based hybridization analysis of human genomic DNA digested with one of the following endonucleases (Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II; DuPont NEN®).

- 5           The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester NY) is exposed to the  
10       blots in a Phosphorimager cassette (Molecular Dynamics, Sunnyvale CA) for several hours, hybridization patterns are compared visually.

#### VII. Antisense Molecules

- An hSBP polypeptide-encoding sequence (which sequences encompass full length and partial hSBP sequences), is used to inhibit in vivo or in vitro expression of naturally occurring  
15       hSBP. Although use of antisense oligonucleotides, comprising about 20 base-pairs, is specifically described, essentially the same procedure is used with larger cDNA fragments. An oligonucleotide based on the coding sequences of hSBP, as shown in Figures 1A and 1B and 2A and 2B is used to inhibit expression of naturally occurring hSBP. The complementary oligonucleotide is designed from the most unique 5' sequence as shown in Figures 1A and 1B and  
20       2A and 2B and used either to inhibit transcription by preventing promoter binding to the upstream nontranslated sequence or translation of an hSBP-encoding transcript by preventing the ribosome from binding. Using an appropriate portion of the leader and 5' sequence of SEQ ID NO:2 or SEQ ID NO:4, an effective antisense oligonucleotide includes any 15-20 nucleotides spanning the region which translates into the signal or early coding sequence of the polypeptide  
25       as shown in Figures 1A and 1B, and 2A and 2B.

#### VIII. Expression of hSBP

- Expression of the hSBP is accomplished by subcloning the cDNAs into appropriate vectors and transfecting the vectors into host cells. In this case, the cloning vector, pSport, previously used for the generation of the cDNA library is used to express hSBP polypeptides in  
30       E. coli. The pSport vector contains a promoter for  $\beta$ -galactosidase upstream of the cloning site, followed by a sequence encoding the amino-terminal Met and the subsequent 7 residues of  $\beta$ -galactosidase. Sequences encoding a bacteriophage promoter useful for transcription and a

linker containing a number of unique restriction sites are positioned immediately after the eight  $\beta$ -galactosidase residue-encoding sequences.

IPTG is used to induce production of the fusion protein in an isolated, transfected bacterial strain according to standard methods. The fusion protein comprises the first seven  
5 residues of  $\beta$ -galactosidase, about 5 to 15 residues of linker, and the full length hSBP-encoding sequence. The signal sequence directs the secretion of hSBP polypeptide into the bacterial growth media, which can then be used directly in the following activity assay.

#### IX. hSBP Activity

Given the homology of hSBP with rat prostatic binding protein (rPBP), human  
10 mammaglobin, rabbit uteroglobin, and FHG 22, activity of hSBP can be assessed by the ability of the polypeptide to bind to steroid. Methods for assessing steroid binding to a polypeptide are well known in the art (see, e.g., Heyns et al. 1977 Eur J Biochem 78:221-230). Alternatively, given the homology between hSBP and rPBP, and the similarities between rPBP and estramucine binding protein (EMBP), hSBP activity can be assessed by the ability of hSBP to bind  
15 estrmucine. Methods for assessing estramucine binding are well known in the art (see, e.g., Appelgren et al. 1979 Acta Pharmacol Toxicol 43:368-374; Försgren et al. 1979 Cancer Res 39:5155-5164; Høisaeter et al. 1981 J Steroid Biochem 14:251-160).

#### X. Production of hSBP Specific Antibodies

hSBP polypeptide substantially purified using PAGE electrophoresis (Sambrook, supra)  
20 is used to immunize rabbits and to produce antibodies using standard protocols. The amino acid sequence translated from hSBP is analyzed using DNASTar software (DNASTar Inc) to determine regions of high immunogenicity, and a corresponding oligopolypeptide is synthesized and used to produce antibodies according to methods known to those of skill in the art. Analysis to select appropriate epitopes, such as those near the C-terminus or in hydrophilic regions is described by  
25 Ausubel et al (supra).

Typically, antibodies are generated using polypeptides about 15 residues in length, which are synthesized on an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry, and coupled to keyhole limpet hemocyanin (KLH, Sigma) by reaction with M-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS; Ausubel et al, supra). Rabbits are  
30 immunized with the polypeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for anti-polypeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radioiodinated.

goat anti-rabbit IgG.

#### **XI. Purification of Naturally Occurring hSBP Using Specific Antibodies**

Naturally-occurring or recombinant hSBP is substantially purified by immunoaffinity chromatography using antibodies specific for hSBP. An immunoaffinity column is constructed  
5 by covalently coupling anti-hSBP antibody to an activated chromatographic resin such as CnBr-activated Sepharose (Pharmacia Biotech). After coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing hSBP polypeptide is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of hSBP (e.g., high  
10 ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody-hSBP binding (e.g., a buffer of pH 2-3 or a high concentration of a chaotrope such as urea or thiocyanate ion), and hSBP polypeptide is collected.

#### **XII. Identification of Molecules Which Interact with hSBP**

hSBP polypeptides, especially biologically active hSBP polypeptides, are labeled with  
15 <sup>125</sup>I Bolton-Hunter reagent (Bolton and Hunter (1973) Biochem J 133:529). Candidate molecules previously arrayed in the wells of a 96 well plate are incubated with the labeled hSBP polypeptides, washed, and assayed for labeled hSBP complex. Data obtained using different concentrations of hSBP are used to calculate values for the number, affinity, and association of hSBP with the candidate molecules.

20 All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited  
25 to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Before the present nucleotide and polypeptide sequences are described, it is to be understood that this invention is not limited to the particular methodology, protocols, cell lines,  
30 vectors and reagents described as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended

claims.

It must be noted that as used herein and in the appended claims, the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells and reference to "the  
5 antibody" includes reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs. Although any methods, devices and materials similar or equivalent to those described  
10 herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now described.

All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are described in the publications which might be used in connection with the presently described invention. The  
15 publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: INCYTE PHARMACEUTICALS, INC.
- (ii) TITLE OF INVENTION: BREAST TUMOR SPECIFIC PROTEINS
- (iii) NUMBER OF SEQUENCES: 13
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
  - (B) STREET: 3174 Porter Drive
  - (C) CITY: Palo Alto
  - (D) STATE: CA
  - (E) COUNTRY: USA
  - (F) ZIP: 94304
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Diskette
  - (B) COMPUTER: IBM Compatible
  - (C) OPERATING SYSTEM: DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) PCT APPLICATION NUMBER: To Be Assigned
  - (B) FILING DATE: Herewith
  - (C) CLASSIFICATION:
- (vii) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/747,547
  - (B) FILING DATE: 12-NOV-1996
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Billings, Lucy J.
  - (B) REGISTRATION NUMBER: 36,749
  - (C) REFERENCE/DOCKET NUMBER: PF-0077 PCT
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (650) 855-0555
  - (B) TELEFAX: (650) 845-4166

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 90 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Lys Leu Ser Val Cys Leu Leu Leu Val Thr Leu Ala Leu Cys Cys  
 1                      5                      10                      15

Tyr	Gln	Ala	Asn	Ala	Glu	Phe	Cys	Pro	Ala	Leu	Val	Ser	Glu	Leu	Leu
		20					25						30		
Asp	Phe	Phe	Phe	Ile	Ser	Glu	Pro	Leu	Phe	Lys	Leu	Ser	Leu	Ala	Lys
		35					40					45			
Phe	Asp	Ala	Pro	Pro	Glu	Ala	Val	Ala	Ala	Lys	Leu	Gly	Val	Lys	Arg
	50					55					60				
Cys	Thr	Asp	Gln	Met	Ser	Leu	Gln	Lys	Arg	Ser	Leu	Ile	Ala	Glu	Val
65					70					75				80	
Leu	Val	Lys	Ile	Leu	Lys	Lys	Cys	Ser	Val						
			85					90							

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

```

(1) (A) LENGTH: 405 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double
    (D) TOPOLOGY: linear

```

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTCCAAATFCA CTCATTGTTT GTGAAAGCTG AGCTCACAGC AAAACAAGCC ACC ATG	56
Met	
1	
AAG CTG TCG GTG TGT CTC CTG CTG GTC ACG CTG GCC CTC TGC TGC TAC	104
Lys Leu Ser Val Cys Leu Leu Leu Val Thr Leu Ala Leu Cys Cys Tyr	
5 10 15	
CAG GCC AAT GCC GAG TTC TGC CCA GCT CTT GTT TCT GAG CTG TTA GAC	152
Gln Ala Asn Ala Glu Phe Cys Pro Ala Leu Val Ser Glu Leu Leu Asp	
20 25 30	
TTC TTC TTC ATT AGT GAA CCT CTG TTC AAG TTA AGT CTT GCC AAA TTT	200
Phe Phe Phe Ile Ser Glu Pro Leu Phe Lys Leu Ser Leu Ala Lys Phe	
35 40 45	
GAT GCC CCT CCG GAA GCT GTT GCA GCC AAG TTA GGA GTG AAG AGA TGC	248
Asp Ala Pro Pro Glu Ala Val Ala Ala Lys Leu Gly Val Lys Arg Cys	
50 55 60 65	
ACG GAT CAG ATG TCC CTT CAG AAA CGA AGC CTC ATT GCG GAA GTC CTG	296
Thr Asp Gln Met Ser Leu Gln Lys Arg Ser Leu Ile Ala Glu Val Leu	
70 75 80	
GTG AAA ATA TTG AAG AAA TGT AGT GTG TGA CATGTAAAAA CTTTCATCCT	346
Val Lys Ile Leu Lys Lys Cys Ser Val *	
85 90	
GGTTTCCACT GTCTTTCAAT GACACCCTGA TCTTCACTGC AGAATGTAAA GGTTC AAC	405

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 93 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Met Lys Leu Leu Met Val Leu Met Leu Ala Ala Leu Ser Gln His Cys
1           5           10           15

Tyr Ala Gly Ser Gly Cys Pro Leu Leu Glu Asn Val Ile Ser Lys Thr
20           25           30

Ile Asn Pro Gln Val Ser Lys Thr Glu Tyr Lys Glu Leu Leu Gln Glu
35           40           45

Phe Ile Asp Asp Asn Ala Thr Thr Asn Ala Ile Asp Glu Leu Lys Glu
50           55           60

Cys Phe Leu Asn Gln Thr Asp Glu Thr Leu Ser Asn Val Glu Val Phe
65           70           75           80

Met Gln Leu Ile Tyr Asp Ser Ser Leu Cys Asp Leu Phe
85           90

```

## (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 495 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

GATCCTTGCC ACCCGCGACT GAACACCGAC AGCAGCAGCC TCACC ATG AAG TTG           54
                                   Met Lys Leu
                                   1

CTG ATG GTC CTC ATG CTG GCG GCC CTC TCC CAG CAC TGC TAC GCA GGC           102
Leu Met Val Leu Met Leu Ala Ala Leu Ser Gln His Cys Tyr Ala Gly
5           10           15

TCT GGC TGC CCC TTA TTG GAG AAT GTG ATT TCC AAG ACA ATC AAT CCA           150
Ser Gly Cys Pro Leu Leu Glu Asn Val Ile Ser Lys Thr Ile Asn Pro
20           25           30           35

CAA GTG TCT AAG ACT GAA TAC AAA GAA CTT CTT CAA GAG TTC ATA GAC           198
Gln Val Ser Lys Thr Glu Tyr Lys Glu Leu Leu Gln Glu Phe Ile Asp
40           45           50

```



GAC AAT GCC ACT ACA AAT GCC ATA GAT GAA TTG AAG GAA TGT TTT CTT	246
Asp Asn Ala Thr Thr Asn Ala Ile Asp Glu Leu Lys Glu Cys Phe Leu	
55 60 65	
AAC CAA ACG GAT GAA ACT CTG AGC AAT GTT GAG GTG TTT ATG CAA TTA	294
Asn Gln Thr Asp Glu Thr Leu Ser Asn Val Glu Val Phe Met Gln Leu	
70 75 80	
ATA TAT GAC ASC AGT CTT TGT GAT TTA TTT TAA CTT TCT GCA AGA CCT	342
Ile Tyr Asp Ser Ser Leu Cys Asp Leu Phe *	
85 90	
TTG GCT CAC AGA ACT GCA GGG TAT GGT GAG AAA CCA ACT ACG GAT TGC	390
TGC AAA CCA CAC CTT CTC TTT CTT ATG TCT TTT TAC TAC AAA CTA CAA	438
GAC AAT TGT TGA AAC CTG CTA TAC ATG TTT ATT TTA ATA AAT TGA TGG	486
CAA AAA CTG	495

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 111 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Ser Thr Ile Lys Leu Ser Leu Cys Leu Leu Ile Met Leu Ala Val	
1 5 10 15	
Cys Cys Tyr Glu Ala Asn Ala Ser Gln Ile Cys Glu Leu Val Ala His	
20 25 30	
Glu Thr Ile Ser Phe Leu Met Lys Ser Glu Glu Glu Leu Lys Lys Glu	
35 40 45	
Leu Glu Met Tyr Asn Ala Pro Pro Ala Ala Val Glu Ala Lys Leu Glu	
50 55 60	
Val Lys Arg Cys Val Asp Gln Met Ser Asn Gly Asp Arg Leu Val Val	
65 70 75 80	
Ala Glu Thr Leu Val Tyr Ile Phe Leu Glu Cys Gly Val Lys Gln Trp	
85 90 95	
Val Glu Thr Tyr Tyr Pro Glu Ile Asp Phe Tyr Tyr Asp Met Asn	
100 105 110	

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 412 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGCTAAGTAG AAAACTGAA ATG AGC ACC ATT AAG CTG AGC CTG TGT CTT CTG	52
Met Ser Thr Ile Lys Leu Ser Leu Cys Leu Leu	
1 5 10	
ATC ATG CTG GCT GTT TGT TGC TAT GAA GCT AAT GCT AGC CAG ATC TGT	100
Ile Met Leu Ala Val Cys Cys Tyr Glu Ala Asn Ala Ser Gln Ile Cys	
15 20 25	
GAA CTT GTT GCC CAT GAA ACC ATA AGC TTC TTA ATG AAA AGT GAG GAA	148
Glu Leu Val Ala His Glu Thr Ile Ser Phe Leu Met Lys Ser Glu Glu	
30 35 40	
GAA CTG AAG AAG GAA CTT GAG ATG TAT AAT GCA CCT CCA GCA GCT GTT	196
Glu Leu Lys Lys Glu Leu Glu Met Tyr Asn Ala Pro Pro Ala Ala Val	
45 50 55	
GAA GCA AAA CTG GAA GTG AAG AGA TGT GTA GAC CAG ATG AGC AAT GGA	244
Glu Ala Lys Leu Glu Val Lys Arg Cys Val Asp Gln Met Ser Asn Gly	
60 65 70 75	
GAC AGA TTG GTA GTA GCA GAA ACA CTG GTA TAC ATT TTT TTG GAA TGT	292
Asp Arg Leu Val Val Ala Glu Thr Leu Val Tyr Ile Phe Leu Glu Cys	
80 85 90	
GGT GTG AAA CAA TGG GTA GAA ACA TAT TAT CCT GAG ATC GAT TTC TAC	340
Gly Val Lys Gln Trp Val Glu Thr Tyr Tyr Pro Glu Ile Asp Phe Tyr	
95 100 105	
TAC GAT ATG AAC TGA TTT TTC CTG TTC AAT GTG ATG GTT TCA AGT CTT	388
Tyr Asp Met Asn *	
110	
GCA CCA ATA AAT TAT TCT CCT TGC	412

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 440 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCTCATCCTT TGCTAAGTCT GAAAAACAAAC TGAGCACCAT GAAGCTGTCC CTGTGTCTTC	60
TGTTGGTCAT CCTGGCTGTT CATTGCTATG TAGCTAATGC TGCAAACGTC TGTCCAGCAG	120
TTCTTTCTGT AAGCAAATCT TTCCTATTTG ACAAGGTGGA GAAATTTGAG GCCTATCTTC	180

```

AGACATTTAA CGCACCTCCA GAGGCTGTTA AAGCAAAAGT GGAAGTGAAG AAATGTATAG      240
ACAGCACTCT GAACTATTTA GAGAAAATGG AAATGGGAAA AATACTGGCA GAAGTCGTTG      300
GGTATTGTAA AGGAACAGAA AACTGAAACA TGGCTCTTCC TGGTCTCCAT TGCTTCTCAC      360
AGATAAACTG ACTTTCCTTG CCCAATGTGA AGGTTECAAC GTCTTGCACT AATAAATTAC      420
TCTCCTTGCA TGTAAAAAAA                                         440

```

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

Met Arg Leu Ser Leu Cys Leu Leu Thr Ile Leu Val Val Cys Cys Tyr
 1             5             10             15
Glu Ala Asn Gly Gln Thr Leu Ala Gly Gln Val Cys Gln Ala Leu Gln
 20             25             30
Asp Val Thr Ile Thr Phe Leu Leu Asn Pro Glu Glu Glu Leu Lys Arg
 35             40             45
Glu Leu Glu Glu Phe Asp Ala Pro Pro Glu Ala Val Glu Ala Asn Leu
 50             55             60
Lys Val Lys Arg Cys Ile Asn Lys Ile Met Tyr Gly Asp Arg Leu Ser
 65             70             75             80
Met Gly Thr Ser Leu Val Phe Ile Met Leu Lys Cys Asp Val Lys Val
 85             90             95
Trp Leu Gln Ile Asn Phe Pro Arg Gly Arg Trp Phe Ser Glu Ile Asn
100            105            110

```

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 91 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```

Met Lys Leu Ala Ile Thr Leu Ala Leu Val Thr Leu Ala Leu Leu Cys
 1             5             10             15

```

Ser Pro Ala Ser Ala Gly Ile Cys Pro Arg Phe Ala His Val Ile Glu  
 20 25 30  
 Asn Leu Leu Leu Gly Thr Pro Ser Ser Tyr Glu Thr Ser Leu Lys Glu  
 35 40 45  
 Phe Glu Pro Asp Asp Thr Met Lys Asp Ala Gly Met Gln Met Lys Lys  
 50 55 60  
 Val Leu Asp Ser Leu Pro Gln Thr Thr Arg Glu Asn Ile Met Lys Leu  
 65 70 75 80  
 Thr Glu Lys Ile Val Lys Ser Pro Leu Cys Met  
 85 90

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 93 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Lys Leu Leu Met Val Leu Met Leu Ala Ala Leu Ser Gln His Cys  
 1 5 10 15  
 Tyr Ala Gly Ser Gly Cys Pro Leu Leu Glu Asn Val Ile Ser Lys Thr  
 20 25 30  
 Ile Asn Pro Gln Val Ser Lys Thr Glu Tyr Lys Glu Leu Leu Gln Glu  
 35 40 45  
 Phe Ile Asp Asp Asn Ala Thr Asn Ala Ile Asp Glu Leu Lys Glu  
 50 55 60  
 Cys Phe Leu Asn Gln Thr Asp Glu Thr Leu Ser Asn Val Glu Val Phe  
 65 70 75 80  
 Met Gln Leu Ile Tyr Asp Ser Ser Leu Cys Asp Leu Phe  
 85 90

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 503 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GACAGCGGCT TCCTTGATCC TTGCCACCCG CGACTGAACA CCGACAGCAG CAGCCTCACC 60

ATG AAG TTG CTG ATG GTC CTC ATG CTG GCG GCC CTC TCC CAG CAC TGC 108  
Met Lys Leu Leu Met Val Leu Met Leu Ala Ala Leu Ser Gln His Cys  
1 5 10 15

TAC GCA GGC TCT GGC TGC CCC TTA TTG GAG AAT GTG ATT TCC AAG ACA 156  
Tyr Ala Gly Ser Gly Cys Pro Leu Leu Glu Asn Val Ile Ser Lys Thr  
20 25 30

ATC AAT CCA CAA GTG TCT AAG ACT GAA TAC AAA GAA CTT CTT CAA GAG 204  
Ile Asn Pro Gln Val Ser Lys Thr Glu Tyr Lys Glu Leu Leu Gln Glu  
35 40 45

TTC ATA GAC GAC AAT GCC ACT ACA AAT GCC ATA GAT GAA TTG AAG GAA 252  
Phe Ile Asp Asp Asn Ala Thr Thr Asn Ala Ile Asp Glu Leu Lys Glu  
50 55 60

TGT TTT CTT AAC CAA ACG GAT GAA ACT CTG AGC AAT GTT GAG GTG TTT 300  
Cys Phe Leu Asn Gln Thr Asp Glu Thr Leu Ser Asn Val Glu Val Phe  
65 70 75 80

ATG CAA TTA ATA TAT GAC AGC AGT CTT TGT GAT TTA TTT TAA CTT TCT 348  
Met Gln Leu Ile Tyr Asp Ser Ser Leu Cys Asp Leu Phe  
85 90

GCA AGA CCT TTG GCT CAC AGA ACT GCA GGG TAT GGT GAG AAA CCA ACT 396

ACG GAT TGC TGC AAA CCA CAC CTT CTC TTT CTT ATG TCT TTT TAC TAC 444

AAA CTA CAA GAC AAT TGT TGA AAC CTG CTA TAC ATG TTT ATT TTA ATA 492

AAT TGA TGG CA 503

## (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 95 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Lys Leu Val Phe Leu Phe Leu Leu Val Thr Ile Pro Ile Cys Cys  
1 5 10 15

Tyr Ala Ser Gly Ser Gly Cys Ser Ile Leu Asp Glu Val Ile Arg Gly  
20 25 30

Thr Ile Asn Ser Thr Val Thr Leu His Asp Tyr Met Lys Leu Val Lys  
35 40 45

Pro Tyr Val Gln Asp His Phe Thr Glu Lys Ala Val Lys Gln Phe Lys  
50 55 60

Gln Cys Phe Leu Asp Gln Thr Asp Lys Thr Leu Glu Asn Val Gly Val  
65 70 75 80

Met Met Glu Ala Ile Phe Asn Ser Glu Ser Cys Gln Gln Pro Ser  
35 90 95

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 509 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGTTTCCTGA TTTCTGTCTT GGACAACAGA ACAACCCACA GGGACTGCCT CAAC ATG 57  
Met  
1

AAG CTG GTG TTT CTA TTC TTG TTG GTC ACC ATC CCT ATT TGC TGC TAT 105  
Lys Leu Val Phe Leu Phe Leu Leu Val Thr Ile Pro Ile Cys Cys Tyr  
5 10 15

GCC AGT GGT TCT GGC TGC AGT ATT CTA GAT GAA GTT ATT AGA GGT ACA 153  
Ala Ser Gly Ser Gly Cys Ser Ile Leu Asp Glu Val Ile Arg Gly Thr  
20 25 30

ATT AAC TCA ACT GTG ACT TTA CAT GAC TAT ATG AAA TTA GTT AAG CCA 201  
Ile Asn Ser Thr Val Thr Leu His Asp Tyr Met Lys Leu Val Lys Pro  
35 40 45

TAT GTA CAA GAT CAT TTT ACT GAA AAG GCT GTG AAG CAA TTC AAG CAG 249  
Tyr Val Gln Asp His Phe Thr Glu Lys Ala Val Lys Gln Phe Lys Gln  
50 55 60 65

TGT TTT CTA GAT CAG ACC GAC AAG ACT CTG GAA AAT GTT GGC GTG ATG 297  
Cys Phe Leu Asp Gln Thr Asp Lys Thr Leu Glu Asn Val Gly Val Met  
70 75 80

ATG GAG GCA ATA TTT AAC AGT GAA AGC TGT CAA CAG CCA TCC TAA ACA 345  
Met Glu Ala Ile Phe Asn Ser Glu Ser Cys Gln Gln Pro Ser  
85 90 95

TCT ACA AGA TCT TTG GCC ACA GGA CTC CAG GAA ACT GGC AAT GGC CAA 393

GCA ACT GAT AAC ACA GAT CAT AAC TCT TCT TTC TTG AAC CCC TTT TTC 441

TAC CTA TAA AGT GCA AGA CGA TTG TTG AAA CCT CAA ATT TAT GTC TTT 489

CCA TTT TAT TAA ATT ATC TG 509

## CLAIMS

1. A substantially purified human steroid binding protein C1 (hSBP1) polypeptide comprising the amino acid sequence of SEQ ID NO:1 or fragments thereof.
2. An isolated and purified polynucleotide sequence encoding an hSBP1 polypeptide of  
5 claim 1.
3. An isolated and purified polynucleotide sequence of claim 2 consisting of SEQ ID NO:2 or variants thereof.
4. A polynucleotide sequence which is complementary to SEQ ID NO:2 or degenerate variants thereof.
- 10 5. A recombinant expression vector comprising the polynucleotide sequence of claim 2.
6. A recombinant host cell containing the polynucleotide sequence of claim 5.
7. A method for producing a polypeptide comprising a polypeptide of SEQ ID NO:1. the method comprising the steps of:
  - a) culturing the host cell of claim 6 under conditions suitable for the expression of the  
15 polypeptide: and
  - b) recovering the polypeptide from the host cell culture.
8. A pharmaceutical composition comprising a substantially purified hSBP polypeptide having an amino acid sequence of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.
- 20 9. A purified antibody that specifically binds the polypeptide of claim 1.
10. A purified antagonist which specifically regulates or modulates the activity of the polypeptide of claim 1.
11. A pharmaceutical composition comprising a substantially purified antagonist of the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.
- 25 12. A substantially purified human steroid binding protein C2 (hSBP2) polypeptide comprising the amino acid sequence of SEQ ID NO:3 or fragments thereof.
13. An isolated and purified polynucleotide sequence encoding an hSBP2 polypeptide of claim 12.
14. An isolated and purified polynucleotide sequence of claim 13 consisting of SEQ ID  
30 NO:4 or variants thereof.
15. A polynucleotide sequence which is complementary to SEQ ID NO:4 or degenerate variants thereof.

16. A recombinant expression vector comprising the polynucleotide sequence of claim 13.
17. A recombinant host cell containing the polynucleotide sequence of claim 13.
18. A method for producing a polypeptide comprising a polypeptide of SEQ ID NO:3, the method comprising the steps of:
  - 5 a) culturing the host cell of claim 17 under conditions suitable for the expression of the polypeptide: and
  - b) recovering the polypeptide from the host cell culture.
19. A pharmaceutical composition comprising a substantially purified human steroid binding protein C2 (hSBP2) polypeptide having an amino acid sequence of SEQ ID NO:3 in conjunction  
10 with a suitable pharmaceutical carrier.
20. A purified antibody that specifically binds the polypeptide of claim 12.
21. A purified antagonist which specifically regulates or modulates the activity of the polypeptide of claim 12.
22. A pharmaceutical composition comprising a substantially purified antagonist of the  
15 polypeptide of claim 12 in conjunction with a suitable pharmaceutical carrier.



1/11

```

5' NGT CCA AAT CAC TCA TTG TTT GTG AAA GCT GAG CTC ACA GCA AAA CAA GCC ACC 54
      9      18      27      36      45
      63      72      81      90      99      108
ATG AAG CTG TCG GTG TGT CTC CTG CTG GTC ACG CTC GCC CTC TGC TGC TAC CAG
M K L S V C L L L L V T L A L C C Y Q
      117      126      135      144      153      162
GCC AAT GCC GAG TTC TGC CCA GCT CTT GTT TCT GAG CTG TTA GAC TTC TTC TTC
A N A E F C P A L V S E L L D F F
      171      180      189      198      207      216
ATT AGT GAA CCT CTG TTC AAG TTA AGT CTT GCC AAA TTT GAT GCC CCT CCG GAA
I S E P L F K L S L A K F D A P P E
      225      234      243      252      261      270
GCT GTT GCA GCC AAG TTA GGA GTG AAG AGA TGC ACG GAT CAG ATG TCC CTT CAG
A V A A K L G V K R C T D Q M S L Q
      279      288      297      306      315      324
AAA CGA AGC CTC ATT GCG GAA GTC CTG GTG AAA ATA TTG AAG AAA TGT AGT GTG
K R S L I A E V L V K I L K K C S V
      333      342      351      360      369      378
TGA CAT GTA AAA ACT TTC ATC CTG GTT TCC ACT GTC TTT CAA TGA CAC CCT GAT
      387      396      405
CTT CAC TGC AGA ATG TAA AGG TTT CAA C 3'

```

FIGURE 1

2/11

```

5' GAT CCT TGC CAC CCG CGA CTG AAC ACC GAC AGC AGC CTC ACC ATG AAG TTG
    9      18      27      36      45      54
    M      K      L

CTG ATG GTC CTC ATG CTG GCG GCC CTC TCC CAG CAC TGC TAC GCA GGC TCT GGC
    63      72      81      90      99      108
    L      M      L      A      A      L      S      Q      H      C      Y      A      G      S      G

TGC CCC TTA TTG GAG AAT GTG ATT TCC AAG ACA ATC AAT CCA CAA GTG TCT AAG
    117      126      135      144      153      162
    C      P      L      L      E      N      V      I      S      K      T      I      N      P      Q      V      S      K

ACT GAA TAC AAA GAA CTT CTT CAA GAG TTC ATA GAC GAC AAT GCC ACT ACA AAT
    171      180      189      198      207      216
    T      E      Y      K      E      L      L      L      Q      E      F      I      D      D      N      A      T      T      N

GCC ATA GAT GAA TTG AAG GAA TGT TTT CTT AAC CAA ACG GAT GAA ACT CTG AGC
    225      234      243      252      261      270
    A      I      D      E      L      K      E      L      E      C      F      L      N      Q      T      D      E      T      L      S

AAT GTT GAG GTG TTT ATG CAA TTA ATA TAT GAC AGC AGT CTT TGT GAT TTA TTT
    279      288      297      306      315      324
    N      V      E      V      F      M      Q      L      I      Y      D      S      S      L      C      D      L      L      F

TAA CTT TCT GCA AGA CCT TTG GCT CAC AGA ACT GCA GGG TAT GGT GAG AAA CCA
    333      342      351      360      369      378

```

FIGURE 2A

3/11

387                    396                    405                    414                    423                    432  
ACT ACG GAT TGC TGC AAA CCA CAC CTT CTC TTT CTT ATG TCT TTT TAC TAC AAA

441                    450                    459                    468                    477                    486  
CTA CAA GAC AAT TGT TGA AAC CTG CTA TAC ATG TTT ATT TTA ATA AAT TGA TGG

495  
CAA AAA CTG 3'

FIGURE 2B

4/11

Library	Lib Description	Abun	Pct Abun
BRSTTUT01	breast tumor, 55 F, match to BRSTNOT02	12	0.1811
THYMNOT02	thymus, 3 M	2	0.0386
BRSTNOT02	breast, 55 F, match to BRSTTUT01	1	0.0192
BRSTNOT01	breast, 56 F	1	0.0158
PROSNOT06	prostate, 57 M, match to PROSTUT04	1	0.0154
BRSTNOT05	breast, 58 F, match to BRSTTUT03	1	0.0148

FIGURE 3

5/11

Library	Lib Description	Abun	Pct Abun
BRSTTUT03	breast tumor, 58 F, match to BRSTNOT05	28	0.4149
BRSTTUT01	breast tumor, 55 F, match to BRSTNOT02	26	0.3923
BRSTNOT02	breast, 55 F, match to BRSTTUT01	5	0.0792
BRSTNOT05	breast, 58 F, match to BRSTTUT03	4	0.0617
BRSTNOT03	breast, 54 F, match to BRSTTUT03	3	0.0440
BRSTNOT01	breast, 56 F	2	0.0384
BRSTNOT04	breast, 62 F	1	0.0319

FIGURE 4

6/11

1	M	-	-	-	K	L	S	V	C	L	L	V	T	L	A	L	C	C	Y	Q	A	N	A	E	-	-	-	-	F	C	P	A	L	V	S	E	L	L	SEQ	ID NO:1		
1	M	S	T	I	K	L	S	L	C	L	L	I	M	-	L	A	V	C	C	Y	E	A	N	-	-	-	-	A	S	Q	I	C	E	L	V	A	H	E	T	I	SEQ	ID NO:5
1	M	-	-	-	R	L	S	L	C	L	L	T	I	-	L	V	V	C	C	Y	E	A	N	G	Q	T	L	A	G	Q	V	C	Q	A	L	Q	D	V	T	I	SEQ	ID NO:8
1	M	-	-	-	K	L	A	I	T	L	A	L	V	T	L	A	L	C	S	P	A	S	A	G	-	-	-	-	I	C	P	R	F	A	H	V	I	E	SEQ	ID NO:9		
33	D	F	F	I	S	E	P	L	F	K	L	S	L	A	K	F	D	A	P	P	E	A	V	A	A	K	L	G	V	K	R	C	T	D	Q	M	S	L	Q	SEQ	ID NO:1	
36	S	F	L	M	K	S	E	E	E	L	K	K	E	L	E	M	Y	N	A	P	P	A	V	E	A	K	L	E	V	K	R	C	V	D	Q	M	S	N	G	SEQ	ID NO:5	
37	T	F	L	L	N	P	E	E	E	L	K	R	E	L	E	E	F	D	A	P	P	E	A	V	E	A	N	L	K	V	K	R	C	I	N	K	I	M	Y	G	SEQ	ID NO:8
33	N	L	L	G	T	P	S	S	Y	E	T	S	L	K	E	F	E	P	D	D	T	M	K	D	A	G	M	Q	M	K	K	V	L	D	S	L	P	Q	T	SEQ	ID NO:9	
73	K	R	S	L	I	A	E	V	L	V	K	I	L	K	K	C	S	V																				SEQ	ID NO:1			
76	D	R	L	V	A	E	T	L	V	Y	I	F	L	E	C	G	V	K	Q	W	V	E	T	Y	Y	P	E	I	D	F	Y	Y	D	M	N					SEQ	ID NO:5	
77	D	R	L	S	M	G	T	S	L	V	F	I	M	L	K	C	D	V	K	V	W	L	Q	I	N	F	P	R	G	R	W	F	S	E	I	N				SEQ	ID NO:8	
73	T	R	E	N	I	M	K	-	L	T	E	K	I	V	K	S	P	L	C	M																		SEQ	ID NO:9			

FIGURE 5

7/11

1	M	K	L	M	V	L	M	L	A	A	L	S	Q	H	C	Y	A	G	S	G	C	P	L	L	E	N	V	I	S	K	T	I	N	P	Q	V	S	K	T	SEQ ID NO:3	
1	M	K	L	L	M	V	L	M	L	A	A	L	S	Q	H	C	Y	A	G	S	G	C	P	L	L	E	N	V	I	S	K	T	I	N	P	Q	V	S	K	T	SEQ ID NO:10
1	M	K	L	V	F	L	F	L	L	V	T	I	P	I	C	C	Y	A	S	G	S	G	C	S	I	L	D	E	V	I	R	G	T	I	N	S	T	V	T	L	SEQ ID NO:12
41	E	Y	K	E	L	L	Q	E	F	I	D	D	N	A	T	T	N	A	I	D	E	L	K	E	C	F	L	N	Q	T	D	E	T	L	S	N	V	E	V	F	SEQ ID NO:3
41	E	Y	K	E	L	L	Q	E	F	I	D	D	N	A	T	T	N	A	I	D	E	L	K	E	C	F	L	N	Q	T	D	E	T	L	S	N	V	E	V	F	SEQ ID NO:10
41	H	D	Y	M	K	L	V	K	P	Y	V	Q	D	H	F	T	E	K	A	V	K	Q	F	K	Q	C	F	L	D	Q	T	D	K	T	L	E	N	V	G	V	SEQ ID NO:12
81	M	Q	L	I	Y	D	S	S	L	C	D	L	F																										SEQ ID NO:3		
81	M	Q	L	I	Y	D	S	S	L	C	D	L	F																									SEQ ID NO:10			
81	M	M	E	A	I	F	N	S	E	S	C	Q	Q	P	S																							SEQ ID NO:12			

FIGURE 6

8/11

1	G	T	C	C	A	A	T	C	A	C	T	C	A	T	T	G	T	G	T	A	A	G	C	T	G	A	G	C	T	C	A	C	A	G	C	SEQ ID NO:2
1	G	-	-	-	-	-	-	-	-	-	C	T	C	A	T	T	T	G	C	T	A	A	G	T	C	T	G	A	-	-	-	-	A	A	C	SEQ ID NO:7
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	G	C	T	A	A	G	T	A	G	A	A	-	-	-	-	A	C	T	SEQ ID NO:6	
41	A	A	A	C	A	A	G	C	C	A	C	C	A	T	G	A	A	G	C	T	G	T	C	C	T	G	T	G	T	C	T	C	T	G	C	SEQ ID NO:2
27	A	A	A	C	T	G	A	G	C	-	A	C	C	A	T	G	A	A	G	C	T	G	T	C	C	T	G	T	C	T	C	T	T	G	C	SEQ ID NO:7
17	G	A	A	T	G	A	G	C	-	A	C	C	A	T	T	A	A	G	C	T	G	A	G	C	C	T	G	T	C	T	T	C	T	-	G	SEQ ID NO:6
81	G	T	C	A	C	G	C	T	G	G	C	C	C	T	C	T	G	C	T	A	C	C	A	G	C	C	A	A	T	G	C	C	G	-	-	SEQ ID NO:2
66	G	T	C	A	T	C	C	T	G	G	C	T	G	T	C	A	T	T	G	C	T	A	T	G	A	A	G	C	T	A	A	T	G	C	A	SEQ ID NO:7
53	A	T	C	A	T	G	C	T	G	G	C	T	G	T	T	T	G	C	T	A	T	G	A	A	G	C	T	A	A	T	G	C	T	A	G	SEQ ID NO:6
118	A	G	T	T	C	T	G	C	C	C	A	G	C	T	C	T	T	T	C	T	G	A	G	C	T	G	T	A	G	A	C	T	T	C	T	SEQ ID NO:2
106	A	C	G	T	C	T	G	T	C	C	A	G	C	A	G	T	T	T	C	T	G	T	A	A	G	C	A	A	T	C	T	T	C	C	T	SEQ ID NO:7
93	A	G	A	T	C	T	G	T	G	A	C	T	T	G	T	T	G	C	C	C	A	T	G	A	A	A	C	A	T	A	A	G	C	T	T	SEQ ID NO:6
158	C	T	T	C	A	T	A	G	T	G	A	A	C	C	T	C	T	T	C	A	A	G	T	T	A	A	G	T	C	T	T	G	C	C	A	SEQ ID NO:2
146	A	T	T	T	G	A	C	A	A	G	T	G	G	A	G	A	T	T	T	G	A	G	C	C	T	A	T	C	T	T	C	A	G	A	C	SEQ ID NO:7
133	A	A	T	G	A	A	A	A	G	T	G	A	G	G	A	A	C	T	T	G	A	A	G	A	A	G	A	A	C	T	T	G	A	G	A	SEQ ID NO:6
198	T	T	T	G	A	T	G	C	C	C	T	C	C	C	T	C	G	G	A	A	G	C	T	G	T	T	G	C	A	A	G	T	A	G	G	SEQ ID NO:2
186	T	T	T	A	A	C	G	C	A	C	C	T	C	C	A	G	A	G	C	T	G	T	T	A	A	A	A	A	G	T	G	G	A	A	G	SEQ ID NO:7
173	T	A	A	T	G	C	A	C	C	T	C	C	A	G	C	T	G	T	T	G	A	A	G	C	A	A	A	A	C	T	G	G	A	A	G	SEQ ID NO:6

FIGURE 7A



9/11

238 T G A A G A G A T G C A C G G A T C A G A T C T C A G A A C G A SEQ ID NO:2  
226 T G A A G A A A T G T A T A G A C A G C A C T A T T A G A G A A SEQ ID NO:7  
213 T G A A G A G A T G T G T A G A C C A G A T G A G C A A T G A G A C A G SEQ ID NO:6

276 A G C C T C A T T G C G G A A G T C T G G T G A A A A T A T T G A A G A A A SEQ ID NO:2  
266 A A T G G A A A T T G G A A A A A A A G A A G T T G G G T A T SEQ ID NO:7  
250 A T T G G T A G T A G C A G A A A C A C T G G T A T A C A T T T T G G A A SEQ ID NO:6

315 T G T A G T G T G T G A C A T - - G T A A A A C T T - - T C A T C C T G G SEQ ID NO:2  
306 T G T A A A G - - G A A C A G A A A A C T G A A A C A T G G C T C T C C T G G SEQ ID NO:7  
290 T G T G G T G A A A C A A T G G G T A G A A A C A T A - - T A T C C T G A SEQ ID NO:6

349 T T C C A C T G T C T T C A A T G A C A C C C T G A T C T T C A C T G C A G SEQ ID NO:2  
344 T C T C C A T T G C T T C T C A C A G A T A A A C T G A C T T T C C T G C C SEQ ID NO:7  
328 G A T C G A T T C T A C T A C G A T A T G A A C T G A T T T C C T G T T C SEQ ID NO:6

389 A A T G T A A A G G T T T C A A C  
384 A A T G T G A A G G T T T C A A C G T C T T G C A C T A A T A A A T T A C T C T SEQ ID NO:2  
368 A A T G T G A T G G T T T C A A G T C T T G C A C C A A T A A A T T A T T C T SEQ ID NO:7

405  
424 C C T T G C A T G T T A A A A A A  
407 C C T T G C SEQ ID NO:6

FIGURE 7B

10/11

```

1  GA - - - - - T C C T T G C C A C C C G C G A C T G A A C A SEQ ID NO:4
1  GA C A G C G G C T T C C T T G A T C C T T G C C A C C C G C G A C T G A A C A SEQ ID NO:11
1  A G - - - - - T T C C T G A T T T C T G A T C T T G G A C A C A G A A C A SEQ ID NO:13

26  - C C G A C A G C A G C A G C C T C A C C C A T G A A G T T G C T G A T G G T C C SEQ ID NO:4
41  - C C G A C A G C A G C A G C C T C A C C C A T G A A G T T G C T G A T G G T C C SEQ ID NO:11
34  A C C C A C A G G G A C T G C C T C A A C A T G A A G C T G G T T C T A T SEQ ID NO:13

65  T C A T G C T G G C G G C C C T C T C C C A G C A C T G C T A C G C - A G - - G SEQ ID NO:4
80  T C A T G C T G G C G G C C C T C T C C C A G C A C T G C T A C G C - A G - - G SEQ ID NO:11
74  T C T T G T T G G T C A C C A T C C T A T T T G C T G C T A T G C C A G T G G SEQ ID NO:13

102 C T C T G G C T G C C C C C T T A T T G G A G A A T G T G A T T C C C A A G A C A SEQ ID NO:4
117 C T C T G G C T G C C C C C T T A T T G G A G A A T G T G A T T T C C C A A G A C A SEQ ID NO:11
114 T T C T G G C T G C C A G T A T T C T A G A T G A A G T A T T A G A G T A C A SEQ ID NO:13

142 A T C A A T C C C A C A A G T G T C T A A G A C T G A A T A C A A A G A A C T T C SEQ ID NO:4
157 A T C A A T C C C A C A A G T G T C T A A G A C T G A A T A C A A A G A A C T T C SEQ ID NO:11
154 A T T A A C T C A A C T G T G A C T T T A C A T G A C T A T G A A A T T A G SEQ ID NO:13

182 T T C A A G A G T T C A T A G A C G A C A A T G C C A C T A C A A A T G C C A T SEQ ID NO:4
197 T T C A A G A G T T C A T A G A C G A C A A T G C C A C T A C A A A T G C C A T SEQ ID NO:11
194 T T A A G C C A T A T G T A C A A G A T C A T T T A C T G A A A A G G C T G T SEQ ID NO:13

222 A G A T G A A T T G A A G G A A T G T T T T C T T A C C C A A A C G G A T G A A SEQ ID NO:4
237 A G A T G A A T T G A A G G A A T G T T T T C T T A A C C C A A A C G G A T G A A SEQ ID NO:11
234 G A A G C A A T T C A A G C A G T G T T T T C T T A G A T C A G A C C G A C A G SEQ ID NO:13

```

FIGURE 8A

11/11

262	A	C	T	C	T	G	A	G	C	A	A	T	G	T	T	G	A	G	G	T	G	T	T	T	A	T	G	C	A	A	T	T	A	A	T	A	T	A	T	A	T	G	SEQ	ID NO:4
277	A	C	T	C	T	G	A	G	C	A	A	T	G	T	T	G	A	G	G	T	G	T	T	T	A	T	G	C	A	A	T	T	A	A	T	A	T	A	T	A	T	G	SEQ	ID NO:11
274	A	C	T	C	T	G	A	A	A	T	G	T	G	G	C	G	T	G	A	T	G	A	T	G	A	T	G	C	A	A	T	G	A	T	A	T	A	T	A	T	G	SEQ	ID NO:13	
302	A	C	A	G	C	-	-	-	A	G	T	C	T	T	T	G	T	G	A	T	T	A	T	T	T	A	A	C	T	T	T	C	T	G	C	A	A	G	SEQ	ID NO:4				
317	A	C	A	G	C	-	-	-	A	G	T	C	T	T	T	G	T	G	A	T	T	A	T	T	T	A	A	C	T	T	T	C	T	G	C	A	A	G	SEQ	ID NO:11				
314	A	C	A	G	T	G	A	A	A	G	C	T	G	T	C	A	A	C	A	G	C	C	A	T	C	C	T	A	A	A	C	A	T	C	T	A	C	A	A	G	SEQ	ID NO:13		
339	A	C	C	T	T	T	G	G	C	T	C	A	C	A	G	A	A	C	T	G	C	A	G	G	T	A	-	T	G	G	T	G	A	G	A	A	C	C	SEQ	ID NO:4				
354	A	C	C	T	T	T	G	G	C	T	C	A	C	A	G	A	A	C	T	G	C	A	G	G	T	A	-	T	G	G	T	G	A	G	A	A	C	C	SEQ	ID NO:11				
354	A	T	C	T	T	T	G	G	C	-	C	A	C	A	G	A	C	T	C	A	G	A	A	A	A	C	A	C	T	G	G	C	A	A	T	G	G	C	C	A	SEQ	ID NO:13		
378	A	A	C	T	A	C	G	G	A	T	T	G	C	T	G	C	A	A	A	C	C	A	C	A	C	C	T	T	C	T	C	T	T	T	C	T	T	A	T	G	SEQ	ID NO:4		
393	A	A	C	T	A	C	G	G	A	T	T	G	C	T	G	C	A	A	A	C	A	C	A	C	A	C	C	T	T	C	T	T	T	C	T	T	A	T	G	SEQ	ID NO:11			
393	A	G	C	A	C	T	G	A	T	A	A	C	-	A	C	A	G	A	T	C	A	T	A	A	C	T	C	T	T	C	T	T	C	T	T	C	T	T	G	A	A	SEQ	ID NO:13	
418	T	C	-	-	-	-	-	-	T	T	T	T	A	C	-	T	A	C	A	A	A	C	T	A	C	A	A	G	A	C	A	A	T	T	G	T	T	G	A	A	A	SEQ	ID NO:4	
433	T	C	-	-	-	-	-	-	T	T	T	T	A	C	-	T	A	C	A	A	A	C	T	A	C	A	A	G	A	C	A	A	T	T	G	T	T	G	A	A	A	SEQ	ID NO:11	
432	C	C	C	T	T	T	T	T	C	T	A	C	C	T	A	C	T	A	A	A	A	G	T	C	A	A	G	A	C	G	A	T	T	G	T	T	G	A	A	A	SEQ	ID NO:13		
453	C	C	T	G	C	T	A	T	A	C	A	T	G	T	T	T	A	T	T	T	A	A	T	A	A	A	T	T	G	A	T	G	G	C	A	A	A	A	A	A	SEQ	ID NO:4		
468	C	C	T	G	C	T	A	T	A	C	A	T	G	T	T	T	A	T	T	T	A	A	A	T	A	A	T	T	G	A	T	G	G	C	A	A	A	A	A	A	SEQ	ID NO:11		
472	C	C	T	C	A	A	A	T	T	A	T	T	A	T	G	T	C	T	-	T	T	C	C	A	T	T	T	A	T	A	T	A	-	-	-	A	T	T	A	T				

FIGURE 8B

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/20674

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C07K14/47 C07K16/18 C12Q1/68 G01N33/574  
G01N33/68 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K C12N C12Q G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WATSON M A ET AL: "MAMMAGLOBIN, A MAMMARY-SPECIFIC MEMBER OF THE UTEROGLOBIN GENE FAMILY, IS OVEREXPRESSED IN HUMAN BREAST CANCER" CANCER RESEARCH, vol. 56, 15 February 1996, pages 860-865, XP002048615 cited in the application	12-18, 20
A	see abstract; figures 1-4 see page 862, paragraph 2 - page 864, paragraph 1 --- -/--	19, 21, 22

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

16 March 1998

Date of mailing of the international search report

26/03/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Oderwald, H

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/20674

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	<p>WO 96 38463 A (UNIV WASHINGTON ; WATSON MARK A (US); FLEMING TIMOTHY P (US)) 5 December 1996 see the claims see abstract; figures 2,4A; examples 1,3,5 see page 3, line 18 - page 5, line 17 see page 19, line 6 - page 21, line 4 ---</p>	12-18,20
A	<p>WATSON M A ET AL: "ISOLATION OF DIFFERENTIALLY EXPRESSED SEQUENCE TAGS FROM HUMAN BREAST CANCER" CANCER RESEARCH, vol. 54, 1 September 1994, pages 4598-4602, XP002019977 see abstract; figures 2,4 see page 4599, paragraph 3 - paragraph 4 ---</p>	
A	<p>PARKER M ET AL: "PROSTATIC STEROID BINDING PROTEIN: GENE DUPLICATION AND STEROID BINDING" NATURE, vol. 298, no. 5869, 1 July 1982, pages 92-94, XP002048614 see abstract; figures 1,2 see page 93, paragraph 3 - page 94, paragraph 3 ---</p>	1-11
A	<p>PARKER M G ET AL.: "Prostatic steroid-binding protein" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 258, no. 1, 10 January 1983, pages 12-15, XP002058665 cited in the application see the whole document -----</p>	12-22

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 97/20674

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9638463 A	05-12-96	US 5668267 A AU 5961696 A	16-09-97 18-12-96
<hr/>			



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : <b>C12N 15/12, C07K 14/47, 16/18, C12Q 1/68, G01N 33/574, 33/68, A61K 38/17</b>		<b>A1</b>	(11) International Publication Number: <b>WO 98/21331</b>
			(43) International Publication Date: <b>22 May 1998 (22.05.98)</b>
(21) International Application Number: <b>PCT/US97/20674</b>		(81) Designated States: AT, AU, BR, CA, CH, CN, DE, DK, ES, FI, GB, IL, JP, KR, MX, NO, NZ, RU, SE, SG, US, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: <b>7 November 1997 (07.11.97)</b>		<p><b>Published</b></p> <p><i>With international search report.</i></p> <p><i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>	
(30) Priority Data: <b>08/747,547 12 November 1996 (12.11.96) US</b>			
(71) Applicant (for all designated States except US): <b>INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).</b>			
(72) Inventors; and (75) Inventors/Applicants (for US only): <b>AKERBLOM, Ingrid, E. [US/US]; 1234 Johnson Street, Redwood City, CA 94061 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). MURRY, Lynn, E. [US/US]; 1124 Los Trancos Road, Portola Valley, CA 94028 (US). GOLI, Surya, K. [IN/US]; 620 Iris Avenue #338, Sunnyvale, CA 94086 (US). HAWKINS, Phillip, R. [US/US]; 750 N. Shoreline Boulevard #96, Mountain View, CA 94034 (US).</b>			
(74) Agent: <b>BILLINGS, Lucy, J.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).</b>			

## (54) Title: HUMAN BREAST TUMOR-SPECIFIC PROTEINS

```

          9      18      27      36      45      54
5'  AGT CCA AAT CAC TCA TTT GTG AAA GCT GAG CTC ACA GCA AAA CAA GCC ACC

          63      72      81      90      99      108
    AGT AAG CTG TCG GTG TGT CTC CTG CTG GTC AGC CTG GCC CTC TGC TGC TAC CAG
    K  K  L  S  V  C  L  L  L  V  T  L  A  L  C  C  Y  Q

          117      126      135      144      153      162
    GGC AAT GCC GAG TTC TGC CCA GCT CTT GTT TCT GAG CTG TTA GAC TTC TTC TTC
    A  E  A  E  F  C  P  A  L  V  S  E  L  L  D  F  F  F

          171      180      189      198      207      216
    ATT AGT GAA GCT CTG TTC AAG TTA AGT CTT GCC AAA TTT GAT GCC GCT CCG GAA
    I  S  E  P  L  F  K  L  S  L  A  K  F  D  A  P  P  E

          225      234      243      252      261      270
    GCT GTT GCA GCC AAG TTA GGA GTG AAG AGA TGC ACU GAT CAG ATG TCC CTT CAG
    A  V  A  A  K  L  G  V  K  R  C  T  D  Q  H  S  L  Q

          279      288      297      306      315      324
    AAA GGA AGC CTC ATT GCG GAA GTC CTG GTG AAA ATA TTT AAG AAA TGT AGT GTG
    K  R  S  L  I  A  E  V  L  V  K  I  L  K  K  C  S  V

          333      342      351      360      369      378
    TGA CAT GTA AAA ACT TTC ATC CTG GTT TCC ACT GTC TTT CAA TGA CAC CCT GAT

          387      396      405
    CTT CAC TGC AGA ATG TAA AGG TTT CAA C 3'

```

## (57) Abstract

The present invention provides polynucleotides that identify and encode two human steroid binding proteins (hSBP). The invention provides for genetically engineered expression vectors and host cells comprising the nucleic acid sequences encoding hSBP polypeptides. The invention also provides for the use of substantially purified hSBP polypeptides, antagonists, and nucleotide sequences (e.g., antisense sequences) in pharmaceutical compositions for the treatment of diseases associated with the expression of hSBP, specifically in the treatment of breast cancer. The invention also describes diagnostic assays for the detection of breast cancer in a susceptible or affected patient. The diagnostic assays utilize compositions comprising the polynucleotides encoding hSBP polypeptides or the complements thereof, which hybridize with the genomic sequence or the transcript of polynucleotides encoding hSBP or anti-hSBP antibodies that specifically bind to an hSBP polypeptide.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						



## HUMAN BREAST TUMOR-SPECIFIC PROTEINS

### TECHNICAL FIELD

The present invention relates to nucleic acid and amino acid sequences of proteins that are differentially expressed in human breast tumor cells and to the use of these sequences in the diagnosis, study, prevention and treatment of disease.

### BACKGROUND ART

Development of breast cancer is associated with multiple genetic changes associated with alterations in expression of specific genes. Breast cancer tissues express genes that are not expressed, or expressed at lower levels, by normal breast tissue. Thus, it is possible to differentiate between normal (non-cancerous) breast tissue and cancerous breast tissue by analyzing differential gene expression between tissues. In addition, there may be several possible alterations that lead to the various possible types of breast cancer. Thus, different types of breast tumors (e.g., invasive vs. non-invasive, ductal vs. axillary lymph node) can be differentiable one from another by the identification of the differences in genes expressed by different types of breast tumor tissues (Porter-Jordan et al. 1994 Hematol Oncol Clin North Am 8:73-100). Breast cancer can thus be generally diagnosed by detection of expression of a gene or genes associated with breast tumor tissue. Where enough information is available about the differential gene expression between various types of breast tumor tissues, the specific type of breast tumor can also be diagnosed.

Nucleotide and amino acid sequences associated with breast tumors can serve as genetic markers of inheritable breast cancer. Genetic changes on chromosome 17 are the most frequently identified events associated with breast tumors. At least four markers on chromosome 17 have been identified: p53 on 17p13.1, regions of loss of heterozygosity (LOH) on 17p13.3 and 17q12-qter, the breast/ovarian cancer locus (BRCA-1) on 17q21, and a fourth breast cancer growth suppressor gene on chromosome 17 (Casey et al. 1993 Hum Molec Genet 2:1921-1927).

Such genetic markers can also be useful in identifying patients susceptible to breast cancer. For example, the genetic marker BRCA-1 has been linked to a susceptibility of developing breast and/or ovarian cancer at a young age in a number of families (Hall et al. 1990 Science 250:1684-1689; Solomon et al. 1991 Cytogenet Cell Genet 58:686-738). The cumulative risks of developing breast cancer associated with the BRCA-1 marker are 50% at 50 years and 82% at 70 years (Easton et al. 1993 Am J Hum Genet 52:678-701). However, since the gene encoding BRCA-1 has not been cloned or sequenced, identification of an individual carrier

of BRCA-1 is not possible without use of linkage analysis. Linkage analysis is generally not feasible in clinical practice since the genetic epidemiology required is tedious, if not impossible, in most cases (Kent et al. 1995 Europ J Surg Oncol 21:240-241).

5 The discovery of nucleotide sequences and polypeptides encoding proteins associated with breast cancer would satisfy a need in the art by providing new means of diagnosing and treating breast cancer.

### DISCLOSURE OF THE INVENTION

10 The present invention features two human steroid binding proteins (hereinafter referred to individually as hSBP1, and hSBP2, and collectively as hSBP), and the full-length nucleotide sequences encoding these proteins, which are differentially expressed in human breast tumor tissue. The transcripts encoding these proteins are present in breast tumor tissue. The first polypeptide, referred to hereinafter as human steroid binding protein C1 (hSBP1), is characterized as having amino acid sequence homology to rat prostatic binding proteins C1 and C2 (PSC1\_RAT and PSC2\_RAT' respectively) and nucleotide sequence homology to hamster  
15 FHG 22 (GI 206441). The second polypeptide, referred to hereinafter as human steroid binding protein C2 (hSBP2), is characterized as having identity to human mammaglobin and homology to rat prostatic binding protein C3 (GI 206448). Accordingly, the invention features two substantially purified human steroid binding proteins, as shown in amino acid sequences of SEQ ID NO:1 and SEQ ID NO:3.

20 One aspect of the invention features isolated and substantially purified polynucleotides that encode hSBP. In a particular aspect, the polynucleotide is the nucleotide sequence of SEQ ID NO:2 and SEQ ID NO:4. In addition, the invention features polynucleotide sequences that hybridize under stringent conditions to SEQ ID NO:2 and SEQ ID NO:4.

25 The invention additionally features nucleic acid sequences encoding hSBP polypeptides, oligonucleotides, peptide nucleic acids (PNA), fragments, portions or antisense molecules thereof, and expression vectors and host cells comprising polynucleotides that encode hSBP. The present invention also relates to antibodies which bind specifically to an hSBP polypeptide, pharmaceutical compositions comprising substantially purified hSBP, fragments thereof, or antagonists of hSBP, in conjunction with a suitable pharmaceutical carrier, and methods for  
30 producing hSBP.

### BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ

ID NO:2) of human steroid binding protein C1, hSBP1. The alignment was produced using MacDNAsis software (Hitachi Software Engineering Co Ltd, San Bruno, CA).

Figures 2A and 2B shows the amino acid sequence (SEQ ID NO:3) and nucleic acid sequence (SEQ ID NO:4) of human steroid binding protein C2, hSBP2 (MacDNAsis software, Hitachi Software Engineering Co Ltd).

Figure 3 shows the northern analysis for the consensus sequence (SEQ ID NO:2) for hSBP1 (Incyte clone 606491). The northern analysis was produced electronically using LIFESEQ™ database (Incyte Pharmaceuticals, Palo Alto CA). The abundance data (Abun) represent the number of transcripts of the gene of interest in the cDNA library. Percent abundance is calculated by dividing the number of transcripts of a gene of interest present in a cDNA library by the total number of transcripts in the cDNA library.

Figure 4 shows the northern analysis for the consensus sequence (SEQ ID NO:4) (LIFESEQ™ database, Incyte Pharmaceuticals, Palo Alto CA).

Figure 5 shows the amino acid sequence alignments among hSBP1 (606491; SEQ ID NO:1) rat prostatic binding proteins C1 and C2 (SEQ ID NOS:5 and 8), and rabbit uteroglobin (SEQ ID NO:9), produced using the multisequence alignment program of DNASTar software (DNASTar Inc, Madison WI).

Figure 6 shows the amino acid sequence alignments among hSBP2 (SEQ ID NO:3) human mammaglobin (GI 1199595; SEQ ID NO:10), and rat prostatic binding protein C3 (GI 206453; SEQ ID NO:12), produced using the multisequence alignment program of DNASTar software (DNASTar Inc, Madison WI).

Figures 7A and 7B shows the nucleotide sequence alignments between hSBP1 (606491; SEQ ID NO:2), hamster FHG22 (GI 1045204; SEQ ID NO:7), and rat prostatic binding protein C1 (GI 206441; SEQ ID NO:6).

Figures 8A and 8B show the nucleotide sequence alignments between hSBP2 (602516; SEQ ID NO:4), human mammaglobin (GI 1199595; SEQ ID NO:11), and rat prostatic binding protein C3 (GI 206452; SEQ ID NO:13).

## MODES FOR CARRYING OUT THE INVENTION

### Definitions

“Nucleic acid sequence” as used herein refers to an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which can be single- or double-stranded, and represent the sense or antisense strand.

Similarly, "amino acid sequence" as used herein refers to an oligopeptide, peptide, polypeptide, or protein sequence. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally-occurring protein molecule, "amino acid sequence" and like terms (e.g., polypeptide, or protein) are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

"Peptide nucleic acid" as used herein refers to a molecule which comprises an oligomer to which an amino acid residue, such as lysine, and an amino group have been added. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary (template) strand of nucleic acid (Nielsen PE et al (1993) Anticancer Drug Des 8:53-63).

As used herein, "SBP" refers to the amino acid sequences of substantially purified steroid binding protein obtained from any species, particularly mammalian, including bovine, ovine, porcine, murine, equine, and preferably human, from any source whether natural, synthetic, semi-synthetic or recombinant. The term "hSBP" as used herein refers to human steroid binding protein and is meant to encompass hSBP1 and hSBP2 polypeptides collectively.

As used herein, "antigenic amino acid sequence" means an amino acid sequence that, either alone or in association with a carrier molecule, can elicit an antibody response in a mammal.

A "variant" of hSBP is defined as an amino acid sequence that is altered by one or more amino acids. The variant can have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant can have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Similar minor variations can also include amino acid deletions or insertions, or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological or immunological activity can be found using computer programs well known in the art, for example, DNASTar software.

A "deletion" is defined as a change in either amino acid or nucleotide sequence in which one or more amino acid or nucleotide residues, respectively, are absent.

An "insertion" or "addition" is that change in an amino acid or nucleotide sequence which has resulted in the addition of one or more amino acid or nucleotide residues, respectively, as compared to the naturally occurring hSBP.

A "substitution" results from the replacement of one or more amino acids or nucleotides

by different amino acids or nucleotides, respectively.

The term "biologically active" refers to a hSBP having structural, regulatory, or biochemical functions of a naturally occurring hSBP. Likewise, "immunologically active" defines the capability of the natural, recombinant or synthetic hSBP, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "derivative" as used herein refers to the chemical modification of a nucleic acid encoding hSBP or the encoded hSBP. Illustrative of such modifications would be replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative would encode a polypeptide which retains essential biological characteristics of natural hSBP.

As used herein, the term "substantially purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated.

"Stringency" typically occurs in a range from about  $T_m - 5^\circ\text{C}$  ( $5^\circ\text{C}$  below the  $T_m$  of the probe) to about  $20^\circ\text{C}$  to  $25^\circ\text{C}$  below  $T_m$ . As will be understood by those of skill in the art, a stringency hybridization can be used to identify or detect identical polynucleotide sequences or to identify or detect similar or related polynucleotide sequences.

The term "hybridization" as used herein shall include "any process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs J (1994) Dictionary of Biotechnology, Stockton Press, New York NY). Amplification as carried out in the polymerase chain reaction technologies is described in Dieffenbach CW and GS Dveksler (1995, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY).

#### **Preferred Embodiments**

The present invention relates to hSBP and to the use of hSBP nucleic acid and amino acid sequences in the study, diagnosis, prevention and treatment of disease. cDNAs encoding a portion of hSBP were predominantly found in cDNA libraries derived from breast tumor tissue (Figures 3 and 4). The abundance data (Abun) reflects the relative level of expression the hSBP sequence in the breast, thymus and prostatic cDNA libraries, with the percentage abundance (Pct Abun) representing the percent of total expressed mRNAs that are homologous to the hSBP sequence.

The present invention also encompasses hSBP variants. A preferred hSBP variant is one

having at least 80% amino acid sequence similarity to an amino acid sequence of an hSBP (i.e., an hSBP1 amino acid sequence (SEQ ID NO:1) or an hSBP2 amino acid sequence (SEQ ID NO:3). A more preferred hSBP variant is one having at least 90% amino acid sequence similarity to SEQ ID NO:1 or SEQ ID NO:3. A most preferred hSBP variant is one having at least 95% amino acid sequence similarity to SEQ ID NO:1 or SEQ ID NO:3.

Nucleic acids encoding the human hSBP of the present invention were first identified in cDNA, Incyte Clones 606491 and 602615 from breast tumor cell cDNA library BRSTTUT01 through a computer-generated search for amino acid sequence alignments. A consensus sequence for each of hSBP1 (SEQ ID NO:2) and hSBP2 (SEQ ID NO: 4) was derived from the overlapping and/or extended nucleic acid sequences as shown in the tables below.

Table 1. Clones from which the consensus sequence (SEQ ID NO:2) of hSBP-C1 was derived.

Sequence I.D.	cDNA Library	Sequence I.D.	cDNA Library	Sequence I.D.	cDNA Library
419412H1	BRSTNOT01	606371H1	BRSTTUT01	1212741H1	BRSTTUT01
603148H1	BRSTTUT01	606491H1	BRSTTUT01	1215122H1	BRSTTUT01
603224H1	BRSTTUT01	825519H1	PROSNOT06	1216374H1	BRSTTUT01
604290H1	BRSTTUT01	967077H1	BRSTNOT05	1217152H1	BRSTTUT01
604954H1	BRSTTUT01	1209955H1	BRSTNOT02		
605120H1	BRSTTUT01	1212005H1	BRSTTUT01		

Table 2. Clones from which the consensus sequence (SEQ ID NO:4) of hSBP-C2 was derived.

Sequence I.D.	cDNA Library	Sequence I.D.	cDNA Library	Sequence I.D.	cDNA Library
410758H1	BRSTNOT01	899784	BRSTTUT03	968163H1	BRSTNOT05
419059H1	BRSTNOT01			977969H1	BRSTNOT02
598065H1	BRSTNOT02	899895H1	BRSTTUT03	10000571H1	BRSTNOT03
601000H1	BRSTNOT02	900118H1	BRSTTUT03	1002776H1	BRSTNOT03
602615H1	BRSTTUT01	901009H1	BRSTTUT03	1004904H1	BRSTNOT02
603548H1	BRSTTUT01	902666H1	BRSTTUT03	1210748H1	BRSTNOT02
603234H1	BRSTTUT01	902354H1	BRSTTUT03		
603999H1	BRSTTUT01	959213H1	BRSTTUT03	1212473H1	BRSTTUT01
605093H1	BRSTTUT01	959506H1	BRSTTUT03	1213350H1	BRSTTUT01
605204H1	BRSTTUT01	960045H1	BRSTTUT03	1213570H1	BRSTTUT01
605215H1	BRSTTUT01	960118H1	BRSTTUT03	1213702H1	BRSTTUT01
605561H1	BRSTTUT01	960656H1	BRSTTUT03	1214253H1	BRSTTUT01
606191H1	BRSTTUT01	962153H1	BRSTTUT03	1214304H1	BRSTTUT01
606289H1	BRSTTUT01	962283H1	BRSTTUT03	1214401H1	BRSTTUT01
606611H1	BRSTTUT01	962488H1	BRSTTUT03	1215366H1	BRSTTUT01
606664H1	BRSTTUT01	962656H1	BRSTTUT03	1215626H1	BRSTTUT01
607089H1	BRSTTUT01	962907H1	BRSTTUT03	1216546H1	BRSTTUT01
897552H1	BRSTNOT05	963043H1	BRSTTUT03	1216653H1	BRSTTUT01
898516H1	BRSTTUT03	963046H1	BRSTTUT03	1216659H1	BRSTTUT01
898821H1	BRSTTUT03	964108H1	BRSTTUT03	1216778H1	BRSTTUT01
899628H1	BRSTTUT03	968127H1	BRSTNOT05		

- 25 The nucleic acid sequence of SEQ ID NO:2 encodes the hSBP1 amino acid sequence, SEQ ID NO:1. The nucleic acid sequence of SEQ ID NO:4 encodes the hSBP2 amino acid sequence. SEQ ID NO:3.

The present invention is based, in part, on the chemical and structural homology between:

1) The amino acid sequence of hSBP1 and rat prostatic binding protein C1 (GI 206442; Delaey et al. 1983 Eur J Biochem 133:645-649) rat prostatic binding protein C2 (Delaey et al. 1987 Nucl Acid Res 15:1627-1641 and rabbit uteroglobin (Menne et al. 1982 Proc Natl Acad Sci USA 79:4853-4857; Figure 5) and the amino acid sequences of hSBP2, human mammaglobin (GI 1100595; SEQ ID NO:10) and rat prostatic binding protein C3 (GI 206453; SEQ ID NO:12;

Figure 6); and 2) The nucleotide sequence encoding hSBP1, rat prostatic binding protein C1 (GI 206442; Delaey et al. supra), and hamster FHG22 (GI 1045204; Dominguez 1995 FEBS Letters 376:257-263; Figures 7A and 7B); and hSBP2, human mammaglobin (GI 1199595; Watson et al. 1996 Cancer Res 56:860-865), and rat prostatic binding protein C3 (GI 206452; Parker et al. 1983 J Biol Chem 258:12-15) (Figures 8A and 8B).

Rat prostatic binding protein (rPBP) is a tetrameric, steroid-binding glycoprotein found in rat ventral prostate, and is the principal protein in rat prostatic fluid (Delaey et al. supra; Parker et al. supra; Heyns et al. 1977 Eur J Biochem 78:221-230; Heyns et al. 1977 Biochem Biophys Res Commun 77:1492-1499; Parker et al. 1978 Eur J Biochem 85:399-406). The rPBP tetramer is composed of two subunits: one subunit containing the polypeptides C1 and C3; and the other subunit containing the polypeptides C2 and C3 (Heyns et al. 1978 Eur J Biochem 89:181-186). rPBP C3 is homologous to human mammaglobin, which in turn is homologous to human Clara cell 10-kilodalton protein and rabbit uteroglobin (Watson et al. supra).

Although rat PBP is primarily expressed in the testes (Lindzey et al. 1994 Vitamins Hormones 49:383-32), transgenic animals harboring a construct containing the 5' flanking region of the rat PBP-C3 gene linked to the coding region for the simian virus 40 large tumor antigen express the transgene in both the prostate and the mammary gland (Allison et al. 1989 Mol Cell Biol 9(5): 2254-2257). The expression of the C3 transgene varies with the sex of the transgenic animal; male transgenic animals express the rat PBP C3 transgene in the prostate and develop prostate carcinoma, while the females express the transgene in the mammary gland and develop atypical mammary hyperplasia (Maroulakou et al. 1994 Proc Natl Acad Sci USA 91:11236-40). Expression of rPBP is regulated by androgenic steroid (e.g., testosterone) partly by stimulating rates of transcription and partly by effects on RNA stability (Parker et al. 1977 Cell 12:401-407; Heyns et al. 1977 Biochem Biophys Res Commun 77:1492-1499; Parker et al. 1979 Proc Natl Acad Sci USA 76:1580-1584; Page et al. 1982 Mol Cell Endocr 27:343-355).

rPBP is similar to estramucine binding protein (EMBP) (Heyns et al. 1977 Eur J Biochem 78:221-30). EMBP is a 46-kDa heterodimer consisting of two closely related subunits, which upon reductive cleavage of disulfide bridges, each subunit is divided into two components. The subunits differ with respect to the components C1 and C2, but share C3 (Bjork et al. 1995 The Prostate (1995) 27:70-83). EMBP binds estramucine (Appelgren et al. 1979 Acta Pharmacol Toxicol 43:368-74; Försgren et al. 1979 Cancer Res 39:5155-64; Høisaeter et al. 1981 J Steroid Biochem 14: 251-60), but does not bind free estrogens (Høisaeter et al. 1981 J Steroid Biochem



14:251-260; Försgren et al. 1979 Proc Natl Acad Sci USA 76:3149-3150). Estramucine, a nitrogen mustard derivative of 17 $\beta$ -estradiol (Mittelman et al. 1977 Cancer Treat Rep 61:307-10; Johnson et al. 1971 Scand J Urol Nephrol 5:103-7), is used to treat patients with prostatic carcinoma. Expression of EMBP is androgen-regulated; this androgen-dependency of EMBP  
5 tends to decline with the transformation of prostatic tissue into biologically more malignant disease (Shiina et al. 1996 Brit J Urol 77:96-101). The ratio of EMBP to dihydroxytestosterone is an indicator of the malignant potential of prostatic carcinoma (Shiina et al. supra).

Rabbit uteroglobin, a homodimeric protein coupled by two disulfide linkages, binds progesterone and structurally related steroids, is also a substrate for transglutaminases, inhibits  
10 phospholipase A<sub>2</sub> activity, and may interfere with the immune and inflammatory activity of several cell types (Miele et al. 1994 J Endocrinol Invest 17:679-692; Miele et al. 1987 Endocrinol Rev 8:474-490). Expression of uteroglobin is regulated by tissue-specific response to steroid hormones (Sandmoller et al. 1994 Oncogene 9:2805-2815).

FHG22 protein was isolated from a female minus male subtracted cDNA library obtained  
15 from the sexually dimorphic Syrian hamster Harderian glands (Dominguez supra). FHG nucleotide and amino acid sequence are similar to the subunits from rat prostatic steroid binding protein C1, uteroglobin (Miele et al. 1994 J Endocrinol Invest 17:679-692), major cat allergen Fel dI (chain I), and mouse salivary androgen binding proteins (subunit  $\alpha$ ) (Karn et al. 1993 Biochem Genet 32:271-277; Dominguez supra). Expression of FHG22 is tissue and sex-  
20 dependent (Dominguez supra).

hSBP1 and rat prostatic binding protein C1 share 55% nucleotide sequence identity at the nucleotide sequence level, whereas hSBP1 and hamster FHG22 share 72% nucleotide sequence identity. hSBP1 is 90 amino acids in length; the amino acid sequence of hSBP1 has 49% identity  
25 with the amino acid sequence of rat prostatic binding protein C1 (SEQ ID NO:5), 44% identity with the amino acid sequence of rat prostatic binding protein C2 (SEQ ID NO:8), and 28% identity with the amino acid sequence of rabbit uteroglobin (SEQ ID NO:9) (Figure 5).

hSBP2 is 93 amino acids in length and shares 99% nucleotide sequence identity with human mammaglobin; the nucleotide sequence of hSBP2 is about 43% identical to the nucleotide sequence of rat prostatic binding protein C3 (Figures 8A and 8B). The amino acid sequence of  
30 hSBP2 is 62% identical to the amino acid sequence of rat prostatic protein C3, and 100% identical to the amino acid sequence of human mammaglobin (Figure 6). Thus, hSBP-C3 is identical to human mammaglobin.

### The hSBP Coding Sequences

The nucleic acid and deduced amino acid sequences of hSBP are shown in Figures 1 (hSBP1) and 2A and 2B (hSBP2). In accordance with the invention, any nucleic acid sequence that encodes an amino acid sequence of an hSBP polypeptide can be used to generate  
5 recombinant molecules which express an hSBP polypeptide. In specific embodiments described herein, a nucleotide sequence encoding a portion of hSBP1 was first isolated as Incyte Clone 606491 from a breast tumor cell line cDNA library BRSTTUT01; and a nucleotide sequence encoding a portion of hSBP2 was first isolated as Incyte Clone 602615 from a breast tumor cell line cDNA library BRSTTUT01.

10 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of degenerate variants of hSBP-encoding nucleotide sequences, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, can be produced. The invention contemplates each and every possible variation of nucleotide sequence that can be made by selecting combinations based on possible codon  
15 choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring hSBP, and all such variations are to be considered as being specifically disclosed herein.

Although nucleotide sequences that encode hSBP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring hSBP under appropriately  
20 selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding hSBP or its derivatives possessing a substantially different codon usage. Codons can be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic expression host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence  
25 encoding hSBP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties (e.g., increased half-life) than transcripts produced from the naturally occurring sequence.

It is now possible to produce a nucleotide sequence encoding an hSBP polypeptide and/or its derivatives entirely by synthetic chemistry, after which the synthetic gene can be inserted into  
30 any of the many available DNA vectors and expression systems using reagents that are well known in the art at the time of the filing of this application. Moreover, synthetic chemistry can be used to introduce mutations into a sequence encoding an hSBP polypeptide.

Also included within the scope of the present invention are polynucleotide sequences that are capable of hybridizing to the nucleotide sequences of Figures 1A-B and/or 2A-B under various conditions of stringency. Hybridization conditions are based on the melting temperature (T<sub>m</sub>) of the nucleic acid binding complex or probe, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA) incorporated herein by reference, and can be used at a defined stringency.

Altered nucleic acid sequences encoding hSBP that can be used in accordance with the invention include deletions, insertions or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent hSBP. The protein can also comprise deletions, insertions or substitutions of amino acid residues that result in a polypeptide that is functionally equivalent to hSBP. Deliberate amino acid substitutions can be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues with the proviso that biological activity of hSBP is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine; glycine, alanine; asparagine, glutamine; serine, threonine, phenylalanine, and tyrosine.

Alleles of hSBP are also encompassed by the present invention. As used herein, an "allele" or "allelic sequence" is an alternative form of hSBP. Alleles result from a mutation (i.e., an alteration in the nucleic acid sequence) and generally produce altered mRNAs and/or polypeptides that may or may not have an altered structure or function relative to naturally-occurring hSBP. Any given gene may have none, one, or many allelic forms. Common mutational changes that give rise to alleles are generally ascribed to natural deletions, additions or substitutions of amino acids. Each of these types of changes may occur alone or in combination with the other changes, and may occur once or multiple times in a given sequence.

Methods for DNA sequencing are well known in the art and employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland OH), Taq polymerase (Perkin Elmer, Norwalk CT), thermostable T7 polymerase (Amersham, Chicago IL), or combinations of recombinant polymerases and proofreading exonucleases such as the ELONGASE Amplification System marketed by Gibco BRL (Gaithersburg MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown MA) and the ABI 377 DNA

sequencers (Perkin Elmer).

### Extending the Polynucleotide Sequence

The polynucleotide sequence encoding hSBP can be extended utilizing partial nucleotide sequence and various methods known in the art to detect upstream sequences such as promoters and regulatory elements. Clones that contain extended sequences are designated by a suffix (see the tables above). Gobinda et al (1993; PCR Methods Applic 2:318-22) disclose "restriction-site" polymerase chain reaction (PCR) as a direct method which uses universal primers to retrieve unknown sequence adjacent to a known locus. First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR can be used to amplify or extend sequences using divergent primers based on a known region (Triglia T et al (1988) Nucleic Acids Res 16:8186). The primers can be designed using OLIGO® 4.06 Primer Analysis Software (1992; National Biosciences Inc, Plymouth MN), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. This method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Capture PCR (Lagerstrom M et al (1991) PCR Methods Applic 1:111-19) is a method for PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. Capture PCR also requires multiple restriction enzyme digestions and ligations to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before PCR.

Another method that can be used to retrieve unknown sequences is that of Parker JD et al (1991; Nucleic Acids Res 19:3055-60). Additionally, one can use PCR, nested primers, and PromoterFinder libraries to "walk in" genomic DNA (PromoterFinder™ Clontech (Palo Alto CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions. Preferably, the libraries used to identify full length cDNAs have been size-selected to include larger cDNAs. More preferably, the cDNA libraries used to identify full-length cDNAs are those generated using random primers, in that such libraries will contain more sequences

comprising regions 5' of the sequence(s) of interest. A randomly primed library can be particularly useful where oligo d(T) libraries do not yield a full-length cDNA. Genomic libraries are preferred for identification and isolation of 5' nontranslated regulatory regions of a sequence(s) of interest.

- 5           Capillary electrophoresis can be used to analyze the size of, or confirm the nucleotide sequence of, sequencing or PCR products. Systems for rapid sequencing are available from Perkin Elmer, Beckman Instruments (Fullerton CA), and other companies. Capillary sequencing can employ flowable polymers for electrophoretic separation, four different, laser-activatable fluorescent dyes (one for each nucleotide), and a charge coupled device camera for detection of
- 10   the wavelengths emitted by the fluorescent dyes. Output/light intensity is converted to electrical signal using appropriate software (e.g. Genotyper™ and Sequence Navigator™ from Perkin Elmer). The entire process from loading of the samples to computer analysis and electronic data display is computer controlled. Capillary electrophoresis is particularly suited to the sequencing of small pieces of DNA that might be present in limited amounts in a particular sample.
- 15   Capillary electrophoresis provides reproducible sequencing of up to 350 bp of M13 phage DNA in 30 min (Ruiz-Martinez MC et al (1993) Anal Chem 65:2851-2858).

#### **Expression of the Nucleotide Sequence**

- In accordance with the present invention, polynucleotide sequences that encode hSBP polypeptides (which polypeptides include fragments of the naturally-occurring polypeptide.
- 20   fusion proteins, and functional equivalents thereof) can be used in recombinant DNA molecules that direct the expression of hSBP in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence, can be used to clone and express hSBP. As will be understood by those of skill in the art, it may be advantageous to produce hSBP-encoding nucleotide
- 25   sequences possessing non-naturally occurring codons. Codons preferred by a particular prokaryotic or eukaryotic host (Murray E et al (1989) Nuc Acids Res 17:477-508) can be selected, for example, to increase the rate of hSBP expression or to produce recombinant RNA transcripts having a desirable characteristic(s) (e.g., longer half-life than transcripts produced from naturally occurring sequence).
- 30   The nucleotide sequences of the present invention can be engineered in order to alter an hSBP coding sequence for a variety of reasons, including but not limited to, alterations that facilitate the cloning, processing and/or expression of the gene product. For example, mutations

can be introduced using techniques that are well known in the art, e.g., site-directed mutagenesis to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, etc.

In another embodiment of the invention, a natural, modified, or recombinant polynucleotide encoding an hSBP polypeptide can be ligated to a heterologous sequence to encode a fusion protein. For example, where an hSBP polypeptide is to be used in a peptide library for screening and identification of inhibitors of hSBP activity, it may be desirable to provide the hSBP polypeptide in the peptide library as a chimeric hSBP protein that can be recognized by a commercially available antibody. A fusion protein can also be engineered to contain a cleavage site located between an hSBP polypeptide-encoding sequence and a heterologous polypeptide sequence, such that the hSBP polypeptide can be cleaved and purified away from the heterologous moiety.

In an alternative embodiment of the invention, a nucleotide sequence encoding an hSBP polypeptide can be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers et al (1980) Nuc Acids Res Symp Ser 215-23, Horn et al (1980) Nuc Acids Res Symp Ser 225-32, etc). Alternatively, the polypeptide itself can be produced using chemical methods to synthesize an hSBP amino acid sequence, in whole or in part. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge et al (1995) Science 269:202-204) and automated synthesis can be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer) in accordance with the instructions provided by the manufacturer.

The newly synthesized peptide can be substantially purified by preparative high performance liquid chromatography (e.g., Creighton (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co, New York NY). The composition of the synthetic peptides can be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally the amino acid sequence of hSBP, or any part thereof, can be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

#### **Expression Systems**

In order to express a biologically active hSBP polypeptide, the nucleotide sequence encoding an hSBP polypeptide or its functional equivalent, is inserted into an appropriate expression vector, i.e., a vector having the necessary elements for the transcription and translation

of the inserted coding sequence.

Methods well known to those skilled in the art can be used to construct expression vectors comprising an hSBP polypeptide-encoding sequence and appropriate transcriptional or translational controls. These methods include in vitro recombinant DNA techniques, synthetic techniques and in vivo recombination or genetic recombination. Such techniques are described in Sambrook et al (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY and Ausubel FM et al (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY.

A variety of expression vector/host systems can be utilized to express an hSBP polypeptide-encoding sequence. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transfected with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with bacterial expression vectors (e.g., Ti or pBR322 plasmid); or animal cell systems.

The "control elements" or "regulatory sequences" of these systems, which vary in their strength and specificities, are those nontranslated regions of the vector, enhancers, promoters, and 3' untranslated regions that interact with host cellular proteins to facilitate transcription and translation of a nucleotide sequence of interest. Depending on the vector system and host utilized, any number of suitable transcriptional and translational elements, including constitutive and inducible promoters, can be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, La Jolla CA) or pSport1 (Gibco BRL), ptrp-lac hybrids, and the like can be used. The baculovirus polyhedron promoter can be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) can be cloned into the vector. In mammalian cell systems, promoters from the mammalian genes or from mammalian viruses are most appropriate. Where it is desirable to generate a cell line containing multiple copies of an hSBP polypeptide-encoding sequence, vectors derived from SV40 or EBV can be used in conjunction with other optional vector elements, e.g., an appropriate selectable marker.

In bacterial systems, a number of expression vectors can be used to express an hSBP polypeptide of interest, and will vary with a variety of factors including the intended use intended

for the hSBP polypeptide produced. For example, when large quantities of an hSBP polypeptide are required (e.g., for the antibody production), vectors that direct high-level expression of fusion proteins that can be readily purified may be desirable. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as Bluescript® (Stratagene; 5 which provides for in-frame ligation of a hSBP polypeptide-encoding sequence with sequences encoding the amino-terminal Met and the subsequent 7 residues of  $\beta$ -galactosidase, thereby producing an hSBP polypeptide- $\beta$ -galactosidase hybrid protein); pIN vectors (Van Heeke & Schuster (1989) J Biol Chem 264:5503-5509); and the like. pGEX vectors (Promega, Madison WI) can also be used to express foreign polypeptides as glutathione S-transferase (GST) fusion 10 proteins. In general, such GST fusion proteins are soluble and can be easily purified from cell lysates by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. GST fusion proteins can be designed to include heparin, thrombin or factor XA protease cleavage sites so that the cloned polypeptide of interest can be readily separated from the GST moiety.

15 Where the host cell is yeast (e.g., *Saccharomyces cerevisiae*) a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH can be used. For reviews, see Ausubel et al (supra) and Grant et al (1987) Methods in Enzymology 153:516-544.

Where plant expression vectors are used, the expression of an hSBP polypeptide-encoding 20 sequence can be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV (Brisson et al (1984) Nature 310:511-514) can be used alone or in combination with the omega leader sequence from TMV (Takamatsu et al (1987) EMBO J 6:307-311). Alternatively, plant promoters, such as the small subunit of RUBISCO (Coruzzi et al (1984) EMBO J 3:1671-1680; Broglie et al (1984) Science 224:838-843) or heat 25 shock promoters (Winter J and Sinibaldi RM (1991) Results Probl Cell Differ 17:85-105), can be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. For reviews of such techniques, see Hobbs S or Murry LE in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill New York NY, pp 191-196 or Weissbach and Weissbach (1988) Methods for Plant Molecular Biology, Academic 30 Press, New York NY, pp 421-463.

Alternatively, insect cell expression systems can be used to express an hSBP polypeptide.. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a



vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The hSBP polypeptide-encoding sequence can be cloned into a nonessential region of the virus, such as the polyhedron gene, and placed under control of the polyhedron promoter. Successful insertion of hSBP renders the polyhedron gene inactive and produces recombinant virus lacking coat protein. The recombinant viruses are then used to infect S. frugiperda cells or Trichoplusia larvae for expression of hSBP polypeptide (Smith et al (1983) J Virol 46:584; Engelhard EK et al (1994) Proc Nat Acad Sci 91:3224-7).

Where the host cell is a mammalian cells, a number of viral-based expression systems can be used. For example, the expression vector can be derived from an adenovirus nucleotide sequence. An hSBP polypeptide-encoding sequence can be ligated into an adenovirus transcription/translation complex, which is composed of the late promoter and tripartite leader sequence. Insertion of the nucleotide sequence of interest into a nonessential E1 or E3 region of the viral genome will result in the production of a viable virus capable of expressing hSBP polypeptide in infected host cells (Logan and Shenk (1984) Proc Natl Acad Sci 81:3655-59). In addition, transcriptional enhancers, such as the Rous sarcoma virus (RSV) enhancer, can be used to increase expression in mammalian host cells.

Specific initiation signals may also be required for efficient translation of an hSBP polypeptide-encoding sequence, e.g., the ATG initiation codon and flanking sequences. Where a native hSBP polypeptide encoding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, where only coding sequence, or a portion thereof, is inserted in an expression vector, exogenous transcriptional control signals including the ATG initiation codon must be provided. Furthermore, the initiation codon must be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be derived from various origins, and can be either natural or synthetic. Expression efficiency can be enhanced by including enhancers appropriate to the cell system in use (Scharf D et al (1994) Results Probl Cell Differ 20:125-62; Bittner et al (1987) Methods in Enzymol 153:516-544).

Host cells can be selected for hSBP polypeptide expression according to the ability of the cell to modulate the expression of the inserted sequences or to process the expressed protein in a desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing that involves cleavage of a "prepro" form of the protein may also be

important for correct polypeptide folding, membrane insertion, and/or function. Host cells such as CHO, HeLa, MDCK, 293, WI38, and others have specific cellular machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing of the introduced, foreign polypeptide.

5           Where long-term, high-yield recombinant polypeptide production is desired, stable expression is preferred. For example, cell lines that stably express hSBP can be transformed using expression vectors containing viral origins of replication or endogenous expression elements and a selectable marker gene. After introduction of the vector, cells can be grown for 1-2 days in an enriched media before they are exposed to selective media. The selectable marker,  
10           which confers resistance to the selective media, allows growth and recovery of cells that successfully express the introduced sequences. Resistant, stably transformed cells can be proliferated using tissue culture techniques appropriate to the host cell type.

          Any number of selection systems can be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler M et al (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy I et al (1980) Cell 22:817-23) genes which can be employed in tk- or apt- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler M et al (1980) Proc Natl Acad Sci 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin F et al (1981) J  
15           Mol Biol 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman SC and RC Mulligan (1988) Proc Natl Acad Sci 85:8047-51). Recently, the use of visible markers has gained popularity with such  
20           markers as anthocyanins,  $\beta$ -glucuronidase and its substrate, GUS, and luciferase and its substrate, luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes CA et al (1995) Methods Mol Biol 55:121-131).

#### **Identification of Transformants Containing the Polynucleotide Sequence**

30           Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression should be confirmed. For example, if the hSBP polypeptide encoding sequence is inserted within a marker gene sequence, recombinant

cells containing this sequence can be identified by the absence of marker gene function.

Alternatively, a marker gene can be placed in tandem with a hSBP sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection is indicative of expression of the tandem hSBP.

5           Alternatively, host cells that contain the coding sequence for hSBP polypeptides and express hSBP polypeptides can be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridization and protein bioassay or immunoassay techniques including membrane, solution, or chip-based technologies for the detection and/or quantitation of the nucleic acid or protein.

10           The presence of the polynucleotide sequence encoding hSBP polypeptides can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes, portions or fragments of polynucleotides encoding hSBP. Nucleic acid amplification-based assays involve the use of oligonucleotides or oligomers based on the hSBP polypeptide-encoding sequence to detect transformants containing hSBP polypeptide-encoding DNA or RNA. As used herein  
15           "oligonucleotides" or "oligomers" refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides which can be used as a probe or amplifier.

          A variety of protocols for detecting and measuring the expression of hSBP, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples  
20           include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on hSBP is preferred, but a competitive binding assay can be employed. These and other assays are described in, e.g., Hampton R et al (1990, Serological Methods, a Laboratory Manual, APS Press, St Paul MN) and Maddox DE et al  
25           (1983, J Exp Med 158:1211).

          A wide variety of detectable labels and conjugation techniques are known by in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to hSBP-encoding polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled  
30           nucleotide. Alternatively, a nucleotide sequence encoding an hSBP polypeptide can be cloned into a vector for the production of an mRNA probe. Such vectors, which are known in the art and commercially available, can be used to synthesize RNA probes in vitro by addition of an

appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides.

A number of companies, including Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH), supply commercial kits and protocols suitable for the methods described above. Suitable reporter molecules or labels include those  
5 radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like, as described in U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241, each of which are incorporated herein by reference. Recombinant immunoglobulins can be produced as according to U.S. Patent No. 4,816,567, incorporated herein by reference.

#### 10 **Purification of hSBP**

Host cells transformed with a nucleotide sequence encoding an hSBP polypeptide can be cultured under conditions suitable for the expression and recovery of the hSBP polypeptide from cell culture. The polypeptide produced by a recombinant cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those  
15 of skill in the art, expression vectors containing polynucleotides encoding hSBP polypeptides can be designed with signal sequences that direct secretion of hSBP through a prokaryotic or eukaryotic cell membrane.

Recombinant hSBP constructs can also include a nucleotide sequence(s) encoding one or more polypeptide domains that, when expressed in-frame with the hSBP-encoding sequence,  
20 facilitates purification of soluble proteins (Kroll DJ et al (1993) DNA Cell Biol 12:441-53; c.f. discussion of vectors infra containing fusion proteins). Such purification facilitating domains include, but are not limited to, metal chelating peptides (e.g., histidine-tryptophan modules) that allow purification with immobilized metals, protein A domains that allow purification with immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity  
25 purification system (Immunex Corp, Seattle WA). A cleavable linker sequences(s) (e.g., Factor XA or enterokinase (Invitrogen, San Diego CA)) between the purification domain and the hSBP polypeptide-encoding sequence can be included to facilitate purification. One such expression vector provides for expression of a fusion protein comprising 6 histidine residues followed by thioredoxin and an enterokinase cleavage site. The histidine residues facilitate purification on  
30 IMIAC (immobilized metal ion affinity chromatography as described in Porath et al (1992) Protein Expression and Purification 3: 263-281), while the enterokinase cleavage site provides a means for separating the hSBP domain from the remainder of the fusion protein.

hSBP polypeptides (which polypeptides encompass polypeptides composed of a portion of the native hSBP amino acid sequence) can also be produced by direct peptide synthesis using solid-phase techniques (cf Stewart et al (1969) Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco; Merrifield J (1963) J Am Chem Soc 85:2149-2154). In vitro protein synthesis  
5 can be performed using manual techniques or by automation. Automated synthesis can be achieved by, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City CA) in accordance with the instructions provided by the manufacturer. Various fragments of hSBP can be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

#### 10 Uses of hSBP

The rationale for use of the nucleotide and polypeptide sequences disclosed herein is based in part on the differential expression of hSBP-encoding sequences in breast tumor tissue and in part on the chemical and structural homology between the hSBP proteins disclosed herein and chemical and structural homology between: 1) hSBP1, rat prostatic binding proteins C1  
15 (GI 206442; Delaey et al. supra), rat prostatic binding protein C2(Delaey et al. 1987 Nucl Acid Res 15:1627-1641) and rabbit uteroglobin (Menne et al. 1982 Proc Natl Acad Sci USA 79:4853-4857) (Figure 5), and 2) hSBP2, human mammaglobin (GI 1199595; Watson et al. supra), and rat prostatic binding protein C3 (GI 206543; Parker et al. supra) (Figure 6).

Accordingly, hSBP or an hSBP derivative can be used in the diagnosis and management  
20 of breast cancer. Given the homology of hSBP with rat PBP, and the differential expression of hSBP in human breast tumor tissue, hSBP can be used as a diagnostic marker for human breast cancer. Expression of rat PBP is regulated by androgens (Muder et al. 1984 Biochem Biophys Acta 781:121-9; Page et al. 1983 Cell 32:495-502) and by growth hormone (Reiter et al. 1995 Endocrinol 166: 3338-44). Thus the level of hSBP can serve as a marker for transformation of  
25 normal breast cells into cancerous cells. Alternatively, or in addition, development of breast cancer can be detected by examining the ratio of hSBP to the levels of steroid hormones (e.g., testosterone or estrogen) or to other hormones (e.g., growth hormone, insulin). Thus expression of hSBP1 and/or hSBP2 can also be used to discriminate between normal and cancerous breast tissue, to discriminate between different types of breast cancer, to provide guidance in selection  
30 of anti-cancer therapies, to monitor the progress of patients undergoing chemotherapy and/or other anti-cancer treatments, to determine the success of surgery to remove cancerous tissue, and to monitor patients who have had or are susceptible to breast cancer. In addition to diagnosis and

treatment of breast cancer after its development. detection of hSBP expression can be used to identify patients susceptible to breast cancer. Expression of hSBP in cancerous cells can be examined in breast tissue in situ or in pathology sections. Alternatively, if hSBP is secreted at sufficient levels, expression of hSBP can be assessed in blood, serum, or plasma. Assessment of levels of hSBP expression can be used to differentiate between normal and cancerous breast tissue, and/or different types of cancerous breast tissue (e.g., invasive vs. non-invasive; ductal vs. axillary lymph node). In addition, because hSBP is differentially expressed in breast tumor cells, hSBP polypeptides can serve as a target for anti-cancer therapy that is targeted to hSBP-expressing breast tumor cells. For example, cells can be transfected with antisense sequences to hSBP-encoding polynucleotides or provided with antagonists to hSBP to reduce or eliminate hSBP expression in cancerous breast cells. Alternatively, cancerous breast cells, or breast cells susceptible to cancer, can be transformed (e.g., via gene therapy techniques) with hSBP-encoding nucleic acid to provide for expression of excess hSBP and interruption of steroid binding.

#### **hSBP Antibodies**

hSBP-specific antibodies are useful for the diagnosis of conditions and diseases associated with expression of hSBP. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library. Neutralizing antibodies, i.e., those which inhibit a biochemical activity of hSBP, are especially preferred for diagnostics and therapeutics.

hSBP polypeptides suitable for production of antibodies need not be biologically active; rather, the polypeptide, or oligopeptide need only be antigenic. Polypeptides used to generate hSBP-specific antibodies generally have an amino acid sequence consisting of at least five amino acids, preferably at least 10 amino acids. Preferably, antigenic hSBP polypeptides mimic an epitope of the native hSBP. Antibodies specific for short hSBP polypeptides can be generated by linking the hSBP polypeptide to a carrier, or fusing the hSBP polypeptide to another protein (e.g., keyhole limpet hemocyanin), and using the carrier-linked or hSBP chimeric molecule as an antigen. In general, anti-hSBP antibodies can be produced according to methods well known in the art.

Various hosts, generally mammalian hosts, can be used to produce anti-hSBP antibodies (e.g., goats, rabbits, rats, mice). Anti-hSBP antibodies are produced by immunizing the host (e.g., by injection) with an hSBP polypeptide that retains immunogenic properties (which encompasses any portion of native hSBP, fragment or oligopeptide). Depending on the host

species, various adjuvants can be used to increase the host's immunological response. Such adjuvants include but are not limited to, Freund's, mineral gels (e.g., aluminum hydroxide), and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are potentially useful human adjuvants.

Monoclonal anti-hSBP antibodies can be prepared using any technique that provides for the production of antibody molecules by immortalized cell lines in culture. These techniques include, but are not limited to, the hybridoma technique originally described by Koehler and Milstein (1975 Nature 256:495-497), the human B-cell hybridoma technique (Kosbor et al (1983) Immunol Today 4:72; Cote et al (1983) Proc Natl Acad Sci 80:2026-2030) and the EBV-hybridoma technique (Cole et al (1985) Monoclonal Antibodies and Cancer Therapy, Alan R Liss Inc, New York NY, pp 77-96).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison et al (1984) Proc Natl Acad Sci 81:6851-6855; Neuberger et al (1984) Nature 312:604-608; Takeda et al (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce hSBP-specific single chain antibodies

Antibodies can be produced in vivo or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi et al (1989, Proc Natl Acad Sci 86: 3833-3837), and Winter G and Milstein C (1991; Nature 349:293-299).

Antibody fragments having specific binding sites for an hSBP polypeptide can also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments, which can be produced by pepsin digestion of the antibody molecule, and Fab fragments, which can be generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries can be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse WD et al (1989) Science 256:1275-1281).

A variety of protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies having established antigen specificities are well known in the art. Such immunoassays typically involve the formation of complexes between an hSBP polypeptide and a specific anti-hSBP antibody, and the detection and quantitation of hSBP-antibody complex formation. A two-site, monoclonal-based immunoassay utilizing monoclonal

antibodies reactive to two noninterfering epitopes on a specific hSBP protein is preferred, but a competitive binding assay can also be employed. These assays are described in Maddox DE et al (1983, J Exp Med 158:1211).

#### **Diagnostic Assays Using hSBP Specific Antibodies**

5           Particular hSBP antibodies are useful for the diagnosis of conditions or diseases characterized by expression of hSBP (e.g., breast cancer) or in assays to monitor patients being treated with hSBP, agonists, antagonists, or inhibitors. Diagnostic assays for hSBP include methods using a detectably-labeled anti-hSBP antibody to detect hSBP in human body fluids or  
10           extracts of cells or tissues. The polypeptides and antibodies of the present invention can be used with or without modification. Frequently, the polypeptides and antibodies are labeled by covalent or noncovalent attachment to a reporter molecule. A wide variety of such suitable reporter molecules are known in the art.

          A variety of protocols for detection and quantifying hSBP, using either polyclonal or monoclonal antibodies specific for an hSBP polypeptide, are known in the art. Examples include  
15           enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on hSBP is preferred, but a competitive binding assay can instead be employed. These assays are described, among other places, in Maddox, DE et al (1983, J Exp Med 158:1211).

20           In order to provide a basis for diagnosis, normal or standard values for hSBP expression must be established. This is accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, preferably human, with antibody to hSBP under conditions suitable for complex formation according to methods well known in the art. The amount of standard complex formation can be quantified by comparing detection levels  
25           associated with known quantities of hSBP with detection levels associated with both control and disease samples from biopsied tissues. Standard values obtained from normal samples are compared with values obtained from samples from subjects potentially affected by disease. Deviation between standard and subject values establishes the presence of disease state.

#### **Drug Screening**

30           hSBP polypeptides, which encompass biologically active or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any of a variety of drug screening techniques. The polypeptide employed in such a test can be free in solution,



affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between hSBP and the agent being tested, can be measured.

Preferably, the drug screening technique used provides for high throughput screening of compounds having suitable binding affinity to the hSBP, as described in detail in "Determination of Amino Acid Sequence Antigenicity" by Geysen HN, WO Application 84/03564, published on September 13, 1984, and incorporated herein by reference. In summary, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with hSBP polypeptides, unreacted materials are washed away, and bound hSBP is detected by methods well known in the art.

Purified hSBP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the polypeptide and immobilize it on a solid support.

The invention also contemplates the use of competitive drug screening assays in which hSBP-specific neutralizing antibodies compete with a test compound for binding of hSBP polypeptide. In this manner, the antibodies can be used to detect the presence of any polypeptide that shares one or more antigenic determinants with an hSBP polypeptide.

#### **Uses of the Polynucleotide Encoding hSBP**

A polynucleotide encoding an hSBP polypeptide (which polypeptides include native hSBP and fragments thereof) can be used for diagnostic and/or therapeutic purposes. For diagnostic purposes, polynucleotides encoding hSBP of this invention can be used to detect and quantitate gene expression in biopsied tissues in which expression of hSBP is implicated, particularly in diagnosis of breast cancer. The diagnostic assay is useful to assess hSBP expression levels (e.g., to distinguish between the absence, and presence of hSBP expression, as well as to assess various hSBP expression levels (e.g., excessively high, high, moderate, or low)) and to monitor regulation of hSBP levels during therapeutic intervention. Included in the scope of the invention are oligonucleotide sequences, antisense RNA and DNA molecules, and peptide nucleic acids (PNAs).

Another aspect of the subject invention is to provide for hybridization or PCR probes capable of detecting polynucleotide sequences encoding hSBP, including genomic sequences and closely related molecules. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' region, and the stringency of the hybridization or amplification (maximal,

high, intermediate or low) will determine whether the probe identifies only naturally occurring sequences encoding hSBP, alleles or related sequences.

The probes of the invention can be used in the detection of related sequences; such probes preferably comprise at least 50% of the nucleotides from any of the hSBP polypeptide-encoding sequences described herein. The hybridization probes of the subject invention can be derived from the nucleotide sequence of SEQ ID NO:2 and SEQ ID NO:4, or from their corresponding genomic sequences including promoters, enhancer elements and introns of the naturally occurring hSBP-encoding sequences. Hybridization probes can be detectably labeled with a variety of reporter molecules, including radionuclides (e.g.,  $^{32}\text{P}$  or  $^{35}\text{S}$ ), or enzymatic labels (e.g., alkaline phosphatase coupled to the probe via avidin/biotin coupling systems), and the like.

Specific hybridization probes for hSBP-encoding DNAs can also be produced by cloning nucleic acid sequences encoding hSBP or hSBP derivatives into vectors for production of mRNA probes. Such vectors, which are known in the art and are commercially available, can be used to synthesize RNA probes *in vitro* using an appropriate RNA polymerase (e.g., T7 or SP6 RNA polymerase) and appropriate radioactively labeled nucleotides.

#### **Diagnostic Use**

Polynucleotide sequences encoding hSBP polypeptide can be used in the diagnosis of conditions or diseases associated with hSBP expression, especially breast cancer. For example, polynucleotide sequences encoding hSBP can be used in hybridization or PCR assays of fluids or tissues from biopsies to detect hSBP expression. Suitable qualitative or quantitative methods include Southern or northern analysis, dot blot or other membrane-based technologies; PCR technologies; dip stick, pIN, chip and ELISA technologies. All of these techniques are well known in the art and are the basis of many commercially available diagnostic kits.

The nucleotide sequences encoding hSBP disclosed herein provide the basis for assays that detect the onset of, susceptibility to, or the presence of breast cancer. Nucleotide sequences encoding hSBP polypeptides can be labeled by methods known in the art and combined with a fluid or tissue sample from a patient suspected of having or susceptible to breast cancer under conditions suitable for the formation of hybridization complexes. After an incubation period, the sample is washed with a compatible fluid which optionally contains a dye (or other label requiring a developer) if the nucleotide has been labeled with an enzyme. After the compatible fluid is rinsed off, the dye is quantitated and compared with a standard. If the amount of dye in the biopsied or extracted sample is significantly elevated over that of a comparable negative

control sample, the nucleotide sequence has hybridized with nucleotide sequences in the sample. The presence of hSBP-encoding nucleotide sequences in the sample, particularly the presence of elevated levels of hSBP-encoding sequences, indicates that the patient has or is at risk of developing the associated disease.

5        Such assays can also be used to evaluate the efficacy of a particular therapeutic treatment regime in animal studies or in clinical trials, or in monitoring the treatment of an individual patient. In order to provide a basis for the diagnosis of disease, a normal or standard profile for hSBP expression must be established. This is accomplished by combining body fluids or cell  
10        extracts taken from normal subjects, either animal or human, with hSBP, or a portion thereof, under conditions suitable for hybridization or amplification. Standard hybridization can be quantified by comparing, in the same experiment, the values obtained for normal subjects with those obtained with a dilution series of hSBP containing known amounts of substantially purified hSBP. Standard values obtained from normal samples are compared with values  
15        obtained from samples from patients afflicted with hSBP-associated diseases, or suspected of having such diseases (e.g., breast cancer). Deviation between standard and subject values is used to establish the presence of disease.

Once disease is established, a therapeutic agent is administered and a treatment profile is generated. Such assays can be repeated on a regular basis to evaluate whether the values in the profile progress toward or return to a normal or standard pattern of hSBP expression. Successive  
20        treatment profiles can be used to show the efficacy of treatment over a period of several days or several months.

Oligonucleotides based upon hSBP sequences can be used in PCR-based techniques, as described in U.S. Patent Nos. 4,683,195 and 4,965,188. Such oligomers are generally chemically synthesized, or produced enzymatically or by recombinantly. Oligomers generally comprise two  
25        nucleotide sequences, one with sense orientation (5'→3') and one with antisense (3'←5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers can be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

30        Additional methods for quantitation of expression of a particular molecule according to the invention include radiolabeling (Melby PC et al 1993 J Immunol Methods 159:235-44) or biotinylating (Duplaa C et al 1993 Anal Biochem 229:36) nucleotides, coamplification of a

control nucleic acid, and interpolation of experimental results according to standard curves. Quantitation of multiple samples can be made more time efficient by running the assay in an ELISA format in which the oligomer of interest is presented in various dilutions and rapid quantitation is accomplished by spectrophotometric or colorimetric detection. For example, the presence of a relatively high amount of hSBP in extracts of biopsied tissues indicates the presence of cancerous breast cells. A definitive diagnosis of this type can allow health professionals to begin aggressive treatment and prevent further worsening of the condition. Similarly, further assays can be used to monitor the progress of a patient during treatment. Furthermore, the nucleotide sequences disclosed herein can be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known such as the triplet genetic code, specific base pair interactions, and the like.

#### Therapeutic Use

Based upon its homology to genes encoding prostatic binding proteins, hSBP polypeptides and its expression profile in breast tumor cells, polynucleotide sequences encoding hSBP disclosed herein may be useful in the treatment of conditions such as breast cancer or other condition associated with hSBP expression or over-expression.

Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids, can be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Recombinant vectors for expression of antisense hSBP polynucleotides can be constructed according to methods well known in the art (see, for example, the techniques described in Sambrook et al (supra) and Ausubel et al (supra)).

Polynucleotides comprising the full length cDNA sequence and/or its regulatory elements enable researchers to use sequences encoding hSBP as an investigative tool in sense (Yousoufian H and HF Lodish 1993 Mol Cell Biol 13:98-104) or antisense (Eguchi et al (1991) Ann Rev Biochem 60:631-652) regulation of gene function. Such technology is now well known in the art, and sense or antisense oligomers, or larger fragments, can be designed from various locations along the coding or control regions.

Expression of genes encoding hSBP can be decreased by transfecting a cell or tissue with expression vectors that express high levels of a desired hSBP-encoding fragment. Such constructs can flood cells with untranslatable sense or antisense sequences. Even in the absence of integration into the DNA, such vectors can continue to transcribe RNA molecules until all

copies are disabled by endogenous nucleases. Transient expression can last for a month or more with a non-replicating vector (Mettler I, personal communication) and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA or PNA, to the control regions of gene encoding hSBP (i.e., the promoters, enhancers, and introns). Oligonucleotides derived from the transcription initiation site, e.g., between -10 and +10 regions of the leader sequence, are preferred. The antisense molecules can also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Similarly, inhibition of expression can be achieved using "triple helix" base-pairing methodology. Triple helix pairing compromises the ability of the double helix to open sufficiently for binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA were reviewed by Gee JE et al (In: Huber BE and BI Carr (1994) Molecular and Immunologic Approaches, Futura Publishing Co, Mt Kisco NY).

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. The invention contemplates engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding hSBP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, which sites include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences between 15 and 20 ribonucleotides corresponding to a region of the target gene containing the cleavage site can be evaluated for secondary structural features that can render the oligonucleotide inoperable. The suitability of candidate targets can also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Antisense molecules and ribozymes of the invention can be prepared by methods known in the art for the synthesis of RNA molecules, including techniques for chemical oligonucleotide synthesis, e.g., solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules can be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding hSBP. Such DNA sequences can be incorporated into a wide variety of vectors with suitable RNA polymerase promoters (e.g, T7 or SP6). Alternatively, antisense cDNA constructs useful in the constitutive or inducible synthesis of antisense RNA can be introduced into cell lines, cells, or tissues.

RNA molecules can be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine and wybutosine as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine that are not as easily recognized by endogenous endonucleases.

Methods for introducing vectors into cells or tissues include those methods discussed *infra* and which are equally suitable for *in vivo*, *in vitro* and *ex vivo* therapy. In *ex vivo* therapy, vectors are introduced into stem cells obtained from the patient and clonally propagated for autologous transplant back into that same patient (see, e.g., U.S. Patent Nos. 5,399,493 and 5,437,994, incorporated herein by reference). Transfection and by liposome methods for delivery of a nucleotide sequence of interest to accomplish gene therapy are well known in the art.

Furthermore, the nucleotide sequences for hSBP disclosed herein can be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including but not limited to such properties as the triplet genetic code and specific base pair interactions.

#### **Detection and Mapping of Related Polynucleotide Sequences**

The hSBP nucleic acid sequences can also be used to generate hybridization probes for mapping the naturally occurring genomic sequence. The sequence can be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. These include *in situ* hybridization to chromosomal spreads, flow-sorted chromosomal preparations, or artificial chromosome constructions such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price CM (1993; Blood Rev 7:127-34) and Trask BJ (1991; Trends Genet 7:149-54).

The technique of fluorescent *in situ* hybridization of chromosome spreads is described in, for example, Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY. Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques can be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a gene encoding hSBP on a physical

chromosomal map and a specific disease (or predisposition to a specific disease) can help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention can be used to detect differences in gene sequences between normal, carrier, or affected individuals.

5        In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers can be used for extending genetic maps. For example an sequence tagged site based map of the human genome was recently published by the Whitehead-MIT Center for Genomic Research (Hudson TJ et al (1995) Science 270:1945-1954). Often the placement of a gene on the chromosome of another mammalian  
10 species such as a mouse (Whitehead Institute/MIT Center for Genome Research, Genetic Map of the Mouse, Database Release 10, April 28, 1995) can reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. Physical mapping provides valuable information to investigators searching for disease genes using positional cloning or other  
15 gene discovery techniques. Once a disease or syndrome, such as ataxia telangiectasia (AT), has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti et al (1988) Nature 336:577-580), other sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequence of the subject invention can also be used to detect differences in the chromosomal location due to  
20 translocation, inversion, etc. among normal, carrier or affected individuals.

#### **Pharmaceutical Compositions**

The present invention relates to pharmaceutical compositions which can comprise nucleotides, proteins, antibodies, agonists, antagonists, or inhibitors, alone or in combination with at least one other agent, such as a stabilizing compound, which can be administered in any  
25 sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. Any of these molecules can be administered to a patient alone or in combination with other agents, drugs or hormones, in pharmaceutical compositions where it is mixed with excipient(s), or with pharmaceutically acceptable carriers. In one embodiment of the present invention, the pharmaceutically acceptable carrier is pharmaceutically inert.

#### **Administration of Pharmaceutical Compositions**

Administration of pharmaceutical compositions is accomplished orally or parenterally. Methods of parenteral delivery include topical, intra-arterial (e.g., directly to the breast tumor),

intramuscular, subcutaneous, intramedullary, intrathecal, intraventricular, intravenous, intraperitoneal, or intranasal administration. In addition to the active ingredients, these pharmaceutical compositions can contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries that facilitate processing of the active compounds into preparations for pharmaceutical use. Further details on techniques for formulation and administration can be found in the latest edition of "Remington's Pharmaceutical Sciences" (Maack Publishing Co, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; and gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores are provided with suitable coatings such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations that can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders such as lactose or starches, lubricants such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycol with or without stabilizers.



Pharmaceutical formulations for parenteral administration include aqueous solutions of active compounds. For injection, the pharmaceutical compositions of the invention can be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection  
5 suspensions can contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds can be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides; or liposomes. Optionally, the suspension can also contain suitable stabilizers or  
10 agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

#### **Manufacture and Storage**

15 The pharmaceutical compositions of the present invention can be manufactured in any suitable manner known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes.

The pharmaceutical composition can be provided as a salt and can be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic,  
20 etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation can be a lyophilized powder in 1mM-50 mM histidine, 0.1%-2% sucrose, 2%-7% mannitol at a pH range of 4.5 to 5.5 that is combined with buffer prior to use.

After pharmaceutical compositions comprising a compound of the invention formulated  
25 in a acceptable carrier have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of hSBP, such labeling would include amount, frequency and method of administration.

#### **Therapeutically Effective Dose**

Pharmaceutical compositions suitable for use in the present invention include  
30 compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model is also used to achieve a desirable concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of protein or its antibodies, antagonists, or inhibitors that ameliorate the symptoms or condition. Therapeutic efficacy and toxicity of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and expressed as the ratio LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The actual dosage can vary within this range depending upon, for example, the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage is chosen by the individual physician in view of the patient to be treated. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Additional factors that may be taken into account include the severity of the disease state, e.g., tumor size and location; age, weight and gender of the patient; diet, time and frequency of administration; drug combination(s); reaction sensitivities; and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

It is contemplated, for example, that hSBP or an hSBP derivative can be delivered in a

suitable formulation to block the progression of breast cancer. Similarly, administration of hSBP antagonists may also inhibit the activity or shorten the lifespan of this protein.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

5

## INDUSTRIAL APPLICABILITY

### I. Construction of BRSTTUT01 cDNA Libraries

The BRSTTUT01 cDNA library was constructed from breast tumor removed from a 55 year old female (lot #0005; Mayo Clinic, Rochester MN). The frozen tissue was immediately homogenized and lysed using a Brinkmann Homogenizer Polytron-PT 3000 (Brinkmann Instruments, Inc. Westbury NY) in guanidinium isothiocyanate solution. Lysates were then loaded on a 5.7 M CsCl cushion and ultracentrifuged in a SW28 swinging bucket rotor for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted once with acid phenol at pH 4.0 and once with phenol chloroform at pH 8.0 and precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in DEPC-treated water and DNase treated for 25 min at 15 37°. The reaction was stopped with an equal volume of acid phenol, and the RNA was isolated using the Qiagen Oligotex kit (QIAGEN Inc, Chatsworth CA) and used to construct the cDNA library. The RNA was handled according to the recommended protocols in the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (catalog #18248-013; Gibco/BRL). cDNAs were fractionated on a Sepharose CL4B column (catalog #275105, Pharmacia), and those 20 cDNAs exceeding 400 bp were ligated into pSport I. The plasmid pSport I was subsequently transformed into DH5a(tm) competent cells (Cat. #18258-012, Gibco/BRL).

### II. Isolation and Sequencing of cDNA Clones From BRSTTUT01

Plasmid DNA was released from the cells and purified using the Miniprep Kit (Catalogue # 77468; Advanced Genetic Technologies Corporation, Gaithersburg MD). This kit consists of a 25 96 well block with reagents for 960 purifications. The recommended protocol was employed except for the following changes: 1) the 96 wells were each filled with only 1 ml of sterile Terrific Broth (Catalog # 22711, LIFE TECHNOLOGIES(tm), Gaithersburg MD) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) the bacteria were cultured for 24 hours after the wells were inoculated and then lysed with 60 µl of lysis buffer; 3) a centrifugation step 30 employing the Beckman GS-6R @2900 rpm for 5 min was performed before the contents of the block were added to the primary filter plate; and 4) the optional step of adding isopropanol to TRIS buffer was not routinely performed. After the last step in the protocol, samples were

transferred to a Beckman 96-well block for storage.

The cDNAs were sequenced by the method of Sanger F and AR Coulson (1975; J Mol Biol 94:441f). using a Hamilton Micro Lab 2200 (Hamilton. Reno NV) in combination with four Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown MA) and Applied Biosystems 377 or 373 DNA Sequencing Systems (Perkin Elmer), and reading frame was determined.

### III. Homology Searching of cDNA Clones and Their Deduced Proteins

Each cDNA was compared to sequences in GenBank using a search algorithm developed by Applied Biosystems and incorporated into the INHERIT™ 670 Sequence Analysis System. In this algorithm, Pattern Specification Language (TRW Inc, Los Angeles CA) was used to determine regions of homology. The three parameters that determine how the sequence comparisons run were window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database was searched for sequences containing regions of homology to the query sequence, and the appropriate sequences were scored with an initial value. Subsequently, these homologous regions were examined using dot matrix homology plots to distinguish regions of homology from chance matches. Smith-Waterman alignments were used to display the results of the homology search.

Peptide and protein sequence homologies were ascertained using the INHERIT- 670 Sequence Analysis System in a way similar to that used in DNA sequence homologies. Pattern Specification Language and parameter windows were used to search protein databases for sequences containing regions of homology which were scored with an initial value. Dot-matrix homology plots were examined to distinguish regions of significant homology from chance matches.

BLAST, which stands for Basic Local Alignment Search Tool (Altschul SF (1993) J Mol Evol 36:290-300; Altschul, SF et al (1990) J Mol Biol 215:403-10), was used to search for local sequence alignments. BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs. BLAST is useful for matches that do not contain gaps. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach identifies HSPs between a query sequence

and a database sequence, evaluates the statistical significance of any matches found, and reports only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output.

#### IV. Northern Analysis

Northern analysis, a laboratory technique used to detect the presence of a gene transcript, and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al. supra).

Analogous computer techniques using BLAST (Altschul SF 1993 and 1990, supra) are used to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ™ database (Incyte, Palo Alto CA). This analysis is much faster than multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores can identify related molecules. The abundance data (Abun) represent the number of transcripts of the gene of interest in the cDNA library. Percent abundance is calculated by dividing the number of transcripts of a gene of interest present in a cDNA library by the total number of transcripts in the cDNA library.

#### V. Extension of hSBP-Encoding Polynucleotides to FullLength or to Recover Regulatory Elements

Full length hSBP-encoding nucleic acid sequences (SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6) are used to design oligonucleotide primers for extending a partial nucleotide sequence to full length and/or for obtaining 5' sequences from genomic libraries. One synthesized primer is used to initiate extension in the antisense direction (XLR), and a second synthesized primer is

used to extend sequence in the sense direction (XLF). Primers allow the extension of the known hSBP-encoding sequence "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest (U.S. Patent Application 08/487,112, filed June 7, 1995, specifically incorporated by reference). The initial primers are designed from the cDNA using  
 5 OLIGO® 4.06 Primer Analysis Software (National Biosciences), or another appropriate program. The initial primers are preferable designed to be 22-30 nucleotides in length, have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. Any stretch of nucleotides that would result in hairpin structures and primer-primer dimerizations is avoided.

10 The original, selected cDNA libraries, or a human genomic library, are used to extend the sequence; the latter is most useful to obtain 5' upstream regions. If more extension is necessary or desired, additional sets of primers are designed to further extend the known region.

By following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix, high fidelity amplification is obtained. Beginning with 40 pmol of  
 15 each primer and the recommended concentrations of all other components of the kit, PCR is performed using the Peltier Thermal Cycler (PTC200; MJ Research, Watertown MA) and the following parameters:

	Step 1	94° C for 1 min (initial denaturation)
	Step 2	65° C for 1 min
20	Step 3	68° C for 6 min
	Step 4	94° C for 15 sec
	Step 5	65° C for 1 min
	Step 6	68° C for 7 min
	Step 7	Repeat step 4-6 for 15 additional cycles
25	Step 8	94° C for 15 sec
	Step 9	65° C for 1 min
	Step 10	68° C for 7:15 min
	Step 11	Repeat step 8-10 for 12 cycles
	Step 12	72° C for 8 min
30	Step 13	4° C (and holding)

A 5-10 µl aliquot of the reaction mixture is analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands containing the largest products were selected and cut out of the  
 35 gel. Further purification is accomplished using a commercial gel extraction method such as QIAQuick™ (QIAGEN Inc). After recovery of the DNA, Klenow enzyme was used to trim single-stranded, nucleotide overhangs creating blunt ends to facilitate religation and cloning.

After ethanol precipitation, the products are redissolved in 13  $\mu$ l of ligation buffer, 1  $\mu$ l T4-DNA ligase (15 units) and 1  $\mu$ l T4 polynucleotide kinase are added, and the mixture is incubated at room temperature for 2-3 hours or overnight at 16° C. Competent *E. coli* cells (in 40  $\mu$ l of appropriate media) are transformed with 3  $\mu$ l of ligation mixture and cultured in 80  $\mu$ l of SOC medium (Sambrook J et al. supra). After incubation for one hour at 37° C, the whole transformation mixture is plated on Luria Bertani (LB)-agar (Sambrook J et al, supra) containing 2xCarb. The following day, several colonies are randomly picked from each plate and cultured in 150  $\mu$ l of liquid LB/2xCarb medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5  $\mu$ l of each overnight culture is transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5  $\mu$ l of each sample was transferred into a PCR array.

For PCR amplification, 18  $\mu$ l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions are run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid and sequenced.

#### VI. Labeling and Use of Hybridization Probes

Hybridization probes derived from SEQ ID NO:2 and SEQ ID NO:4 are used to screen cDNAs, genomic DNAs or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base-pairs, is specifically described, essentially the same procedure is used with larger cDNA fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences), labeled by combining 50 pmol of each oligomer and 250 mCi of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham, Chicago IL) and T4 polynucleotide kinase (DuPont NEN®, Boston MA). The labeled oligonucleotides are substantially purified with Sephadex G-25

super fine resin column (Pharmacia). A portion containing  $10^7$  counts per minute of each of the sense and antisense oligonucleotides is used in a typical membrane based hybridization analysis of human genomic DNA digested with one of the following endonucleases (Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II; DuPont NEN®).

5 The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester NY) is exposed to the  
10 blots in a Phosphorimager cassette (Molecular Dynamics, Sunnyvale CA) for several hours, hybridization patterns are compared visually.

#### VII. Antisense Molecules

An hSBP polypeptide-encoding sequence (which sequences encompass full length and partial hSBP sequences), is used to inhibit in vivo or in vitro expression of naturally occurring  
15 hSBP. Although use of antisense oligonucleotides, comprising about 20 base-pairs, is specifically described, essentially the same procedure is used with larger cDNA fragments. An oligonucleotide based on the coding sequences of hSBP, as shown in Figures 1A and 1B and 2A and 2B is used to inhibit expression of naturally occurring hSBP. The complementary oligonucleotide is designed from the most unique 5' sequence as shown in Figures 1A and 1B and  
20 2A and 2B and used either to inhibit transcription by preventing promoter binding to the upstream nontranslated sequence or translation of an hSBP-encoding transcript by preventing the ribosome from binding. Using an appropriate portion of the leader and 5' sequence of SEQ ID NO:2 or SEQ ID NO:4, an effective antisense oligonucleotide includes any 15-20 nucleotides spanning the region which translates into the signal or early coding sequence of the polypeptide  
25 as shown in Figures 1A and 1B, and 2A and 2B.

#### VIII. Expression of hSBP

Expression of the hSBP is accomplished by subcloning the cDNAs into appropriate vectors and transfecting the vectors into host cells. In this case, the cloning vector, pSport, previously used for the generation of the cDNA library is used to express hSBP polypeptides in  
30 E. coli. The pSport vector contains a promoter for  $\beta$ -galactosidase upstream of the cloning site, followed by a sequence encoding the amino-terminal Met and the subsequent 7 residues of  $\beta$ -galactosidase. Sequences encoding a bacteriophage promoter useful for transcription and a



linker containing a number of unique restriction sites are positioned immediately after the eight  $\beta$ -galactosidase residue-encoding sequences.

IPTG is used to induce production of the fusion protein in an isolated, transfected bacterial strain according to standard methods. The fusion protein comprises the first seven residues of  $\beta$ -galactosidase, about 5 to 15 residues of linker, and the full length hSBP-encoding sequence. The signal sequence directs the secretion of hSBP polypeptide into the bacterial growth media, which can then be used directly in the following activity assay.

#### IX. hSBP Activity

Given the homology of hSBP with rat prostatic binding protein (rPBP), human mammaglobin, rabbit uteroglobin, and FHG 22, activity of hSBP can be assessed by the ability of the polypeptide to bind to steroid. Methods for assessing steroid binding to a polypeptide are well known in the art (see, e.g., Heyns et al. 1977 *Eur J Biochem* 78:221-230). Alternatively, given the homology between hSBP and rPBP, and the similarities between rPBP and estramucine binding protein (EMBP), hSBP activity can be assessed by the ability of hSBP to bind estramucine. Methods for assessing estramucine binding are well known in the art (see, e.g., Appelgren et al. 1979 *Acta Pharmacol Toxicol* 43:368-374; Försgren et al. 1979 *Cancer Res* 39:5155-5164; Høisaeter et al. 1981 *J Steroid Biochem* 14:251-160).

#### X. Production of hSBP Specific Antibodies

hSBP polypeptide substantially purified using PAGE electrophoresis (Sambrook, supra) is used to immunize rabbits and to produce antibodies using standard protocols. The amino acid sequence translated from hSBP is analyzed using DNASTar software (DNASTar Inc) to determine regions of high immunogenicity, and a corresponding oligopolypeptide is synthesized and used to produce antibodies according to methods known to those of skill in the art. Analysis to select appropriate epitopes, such as those near the C-terminus or in hydrophilic regions is described by Ausubel et al (supra).

Typically, antibodies are generated using polypeptides about 15 residues in length, which are synthesized on an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry, and coupled to keyhole limpet hemocyanin (KLH, Sigma) by reaction with M-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS; Ausubel et al, supra). Rabbits are immunized with the polypeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for anti-polypeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radioiodinated.

goat anti-rabbit IgG.

#### **XI. Purification of Naturally Occurring hSBP Using Specific Antibodies**

Naturally-occurring or recombinant hSBP is substantially purified by immunoaffinity chromatography using antibodies specific for hSBP. An immunoaffinity column is constructed  
5 by covalently coupling anti-hSBP antibody to an activated chromatographic resin such as CnBr-activated Sepharose (Pharmacia Biotech). After coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing hSBP polypeptide is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of hSBP (e.g., high  
10 ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody-hSBP binding (e.g., a buffer of pH 2-3 or a high concentration of a chaotrope such as urea or thiocyanate ion), and hSBP polypeptide is collected.

#### **XII. Identification of Molecules Which Interact with hSBP**

hSBP polypeptides, especially biologically active hSBP polypeptides, are labeled with  
15 <sup>125</sup>I Bolton-Hunter reagent (Bolton and Hunter (1973) Biochem J 133:529). Candidate molecules previously arrayed in the wells of a 96 well plate are incubated with the labeled hSBP polypeptides, washed, and assayed for labeled hSBP complex. Data obtained using different concentrations of hSBP are used to calculate values for the number, affinity, and association of hSBP with the candidate molecules.

20 All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited  
25 to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Before the present nucleotide and polypeptide sequences are described, it is to be understood that this invention is not limited to the particular methodology, protocols, cell lines,  
30 vectors and reagents described as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended

claims.

It must be noted that as used herein and in the appended claims, the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells and reference to "the antibody" includes reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs. Although any methods, devices and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now described.

All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are described in the publications which might be used in connection with the presently described invention. The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: INCYTE PHARMACEUTICALS, INC.
- (ii) TITLE OF INVENTION: BREAST TUMOR SPECIFIC PROTEINS
- (iii) NUMBER OF SEQUENCES: 13
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
  - (B) STREET: 3174 Porter Drive
  - (C) CITY: Palo Alto
  - (D) STATE: CA
  - (E) COUNTRY: USA
  - (F) ZIP: 94304
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Diskette
  - (B) COMPUTER: IBM Compatible
  - (C) OPERATING SYSTEM: DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) PCT APPLICATION NUMBER: To Be Assigned
  - (B) FILING DATE: Herewith
  - (C) CLASSIFICATION:
- (vii) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/747,547
  - (B) FILING DATE: 12-NOV-1996
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Billings, Lucy J.
  - (B) REGISTRATION NUMBER: 36,749
  - (C) REFERENCE/DOCKET NUMBER: PF-0077 PCT
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (650) 855-0555
  - (B) TELEFAX: (650) 845-4166

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 90 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:  

Met	Lys	Leu	Ser	Val	Cys	Leu	Leu	Leu	Val	Thr	Leu	Ala	Leu	Cys	Cys
1				5					10					15	

Tyr Gln Ala Asn Ala Glu Phe Cys Pro Ala Leu Val Ser Glu Leu Leu  
 20 25 30  
 Asp Phe Phe Phe Ile Ser Glu Pro Leu Phe Lys Leu Ser Leu Ala Lys  
 35 40 45  
 Phe Asp Ala Pro Pro Glu Ala Val Ala Ala Lys Leu Gly Val Lys Arg  
 50 55 60  
 Cys Thr Asp Gln Met Ser Leu Gln Lys Arg Ser Leu Ile Ala Glu Val  
 65 70 75 80  
 Leu Val Lys Ile Leu Lys Lys Cys Ser Val  
 85 90

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 405 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTCCAAATCA CTCATTGTTT GTGAAAGCTG AGCTCACAGC AAAACAAGCC ACC ATG 56  
 Met  
 1  
 AAG CTG TCG GTG TGT CTC CTG CTG GTC ACG CTG GCC CTC TGC TGC TAC 104  
 Lys Leu Ser Val Cys Leu Leu Leu Val Thr Leu Ala Leu Cys Cys Tyr  
 5 10 15  
 CAG GCC AAT GCC GAG TTC TGC CCA GCT CTT GTT TCT GAG CTG TTA GAC 152  
 Gln Ala Asn Ala Glu Phe Cys Pro Ala Leu Val Ser Glu Leu Leu Asp  
 20 25 30  
 TTC TTC TTC ATT AGT GAA CCT CTG TTC AAG TTA AGT CTT GCC AAA TTT 200  
 Phe Phe Phe Ile Ser Glu Pro Leu Phe Lys Leu Ser Leu Ala Lys Phe  
 35 40 45  
 GAT GCC CCT CCG GAA GCT GTT GCA GCC AAG TTA GGA GTG AAG AGA TGC 248  
 Asp Ala Pro Pro Glu Ala Val Ala Ala Lys Leu Gly Val Lys Arg Cys  
 50 55 60 65  
 ACG GAT CAG ATG TCC CTT CAG AAA CGA AGC CTC ATT GCG GAA GTC CTG 296  
 Thr Asp Gln Met Ser Leu Gln Lys Arg Ser Leu Ile Ala Glu Val Leu  
 70 75 80  
 GTG AAA ATA TTG AAG AAA TGT AGT GTG TGA CATGTAAAAA CTTTCATCCT 346  
 Val Lys Ile Leu Lys Lys Cys Ser Val \*  
 85 90  
 GGTTCCTACT GTCTTTCAAT GACACCCTGA TCTTCACTGC AGAATGTAAA GGTTC AAC 405

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 93 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Met Lys Leu Leu Met Val Leu Met Leu Ala Ala Leu Ser Gln His Cys
 1             5             10             15
Tyr Ala Gly Ser Gly Cys Pro Leu Leu Glu Asn Val Ile Ser Lys Thr
 20             25             30
Ile Asn Pro Gln Val Ser Lys Thr Glu Tyr Lys Glu Leu Leu Gln Glu
 35             40             45
Phe Ile Asp Asp Asn Ala Thr Thr Asn Ala Ile Asp Glu Leu Lys Glu
 50             55             60
Cys Phe Leu Asn Gln Thr Asp Glu Thr Leu Ser Asn Val Glu Val Phe
 65             70             75             80
Met Gln Leu Ile Tyr Asp Ser Ser Leu Cys Asp Leu Phe
 85             90

```

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 495 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

GATCCTTGCC ACCCGCGACT GAACACCGAC AGCAGCAGCC TCACC ATG AAG TTG      54
                               Met Lys Leu
                               1
CTG ATG GTC CTC ATG CTG GCG GCC CTC TCC CAG CAC TGC TAC GCA GGC      102
Leu Met Val Leu Met Leu Ala Leu Ser Gln His Cys Tyr Ala Gly
 5             10             15
TCT GGC TGC CCC TTA TTG GAG AAT GTG ATT TCC AAG ACA ATC AAT CCA      150
Ser Gly Cys Pro Leu Leu Glu Asn Val Ile Ser Lys Thr Ile Asn Pro
 20             25             30             35
CAA GTG TCT AAG ACT GAA TAC AAA GAA CTT CTT CAA GAG TTC ATA GAC      198
Gln Val Ser Lys Thr Glu Tyr Lys Glu Leu Leu Gln Glu Phe Ile Asp
 40             45             50

```

GAC AAT GCC ACT ACA AAT GCC ATA GAT GAA TTG AAG GAA TGT TTT CTT	246
Asp Asn Ala Thr Thr Asn Ala Ile Asp Glu Leu Lys Glu Cys Phe L u	
55 60 65	
AAC CAA ACG GAT GAA ACT CTG AGC AAT GTT GAG GTG TTT ATG CAA TTA	294
Asn Gln Thr Asp Glu Thr Leu Ser Asn Val Glu Val Phe Met Gln Leu	
70 75 80	
ATA TAT GAC AGC AGT CTT TGT GAT TTA TTT TAA CTT TCT GCA AGA CCT	342
Ile Tyr Asp Ser Ser Leu Cys Asp Leu Phe *	
85 90	
TTG GCT CAC AGA ACT GCA GGG TAT GGT GAG AAA CCA ACT ACG GAT TGC	390
TGC AAA CCA CAC CTT CTC TTT CTT ATG TCT TTT TAC TAC AAA CTA CAA	438
GAC AAT TGT TGA AAC CTG CTA TAC ATG TTT ATT TTA ATA AAT TGA TGG	486
CAA AAA CTG	495

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 111 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Ser Thr Ile Lys Leu Ser Leu Cys Leu Leu Ile Met Leu Ala Val	
1 5 10 15	
Cys Cys Tyr Glu Ala Asn Ala Ser Gln Ile Cys Glu Leu Val Ala His	
20 25 30	
Glu Thr Ile Ser Phe Leu Met Lys Ser Glu Glu Glu Leu Lys Lys Glu	
35 40 45	
Leu Glu Met Tyr Asn Ala Pro Pro Ala Ala Val Glu Ala Lys Leu Glu	
50 55 60	
Val Lys Arg Cys Val Asp Gln Met Ser Asn Gly Asp Arg Leu Val Val	
65 70 75 80	
Ala Glu Thr Leu Val Tyr Ile Phe Leu Glu Cys Gly Val Lys Gln Trp	
85 90 95	
Val Glu Thr Tyr Tyr Pro Glu Ile Asp Phe Tyr Tyr Asp Met Asn	
100 105 110	

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 412 base pairs

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGCTAAGTAG AAAACTGAA ATG AGC ACC ATT AAG CTG AGC CTG TGT CTT CTG	52
Met Ser Thr Ile Lys Leu Ser Leu Cys Leu Leu	
1 5 10	
ATC ATG CTG GCT GTT TGT TGC TAT GAA GCT AAT GCT AGC CAG ATC TGT	100
Ile Met Leu Ala Val Cys Cys Tyr Glu Ala Asn Ala Ser Gln Ile Cys	
15 20 25	
GAA CTT GTT GCC CAT GAA ACC ATA AGC TTC TTA ATG AAA AGT GAG GAA	148
Glu Leu Val Ala His Glu Thr Ile Ser Phe Leu Met Lys Ser Glu Glu	
30 35 40	
GAA CTG AAG AAG GAA CTT GAG ATG TAT AAT GCA CCT CCA GCA GCT GTT	196
Glu Leu Lys Lys Glu Leu Glu Met Tyr Asn Ala Pro Pro Ala Ala Val	
45 50 55	
GAA GCA AAA CTG GAA GTG AAG AGA TGT GTA GAC CAG ATG AGC AAT GGA	244
Glu Ala Lys Leu Glu Val Lys Arg Cys Val Asp Gln Met Ser Asn Gly	
60 65 70 75	
GAC AGA TTG GTA GTA GCA GAA ACA CTG GTA TAC ATT TTT TTG GAA TGT	292
Asp Arg Leu Val Val Ala Glu Thr Leu Val Tyr Ile Phe Leu Glu Cys	
80 85 90	
GGT GTG AAA CAA TGG GTA GAA ACA TAT TAT CCT GAG ATC GAT TTC TAC	340
Gly Val Lys Gln Trp Val Glu Thr Tyr Tyr Pro Glu Ile Asp Phe Tyr	
95 100 105	
TAC GAT ATG AAC TGA TTT TTC CTG TTC AAT GTG ATG GTT TCA AGT CTT	388
Tyr Asp Met Asn *	
110	
GCA CCA ATA AAT TAT TCT CCT TGC	412

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 440 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCTCATCCTT TGCTAAGTCT GAAAACAAAC TGAGCACCAT GAAGCTGTCC CTGTGTCTTC	60
TGTTGGTCAT CCTGGCTGTT CATTGCTATG AAGCTAATGC TGCAAACGTC TGTCCAGCAG	120
TTCTTTCTGT AAGCAAATCT TTCCTATTTG ACAAGGTGGA GAAATTTGAG GCCTATCTTC	180



```

AGACATTTAA CGCACCTCCA GAGGCTGTTA AAGCAAAGT GGAAGTGAAG AAATGTATAG      240
ACAGCACTCT GAACTATTTA GAGAAAATGG AAATGGGAAA AATACTGGCA GAAGTCGTTG      300
GGTATTGTAA AGGAACAGAA AACTGAAACA TGGCTCTTCC TGGTCTCCAT TGCTTCTCAC      360
AGATAAACTG ACTTTCCTTG CCCAATGTGA AGGTTTCAAC GTCTTGCACT AATAAATTAC      420
TCTCCTTGCA TGTAAAAAAA                                         440

```

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 112 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

Met Arg Leu Ser Leu Cys Leu Leu Thr Ile Leu Val Val Cys Cys Tyr
1      5      10      15
Glu Ala Asn Gly Gln Thr Leu Ala Gly Gln Val Cys Gln Ala Leu Gln
20      25      30
Asp Val Thr Ile Thr Phe Leu Leu Asn Pro Glu Glu Glu Leu Lys Arg
35      40      45
Glu Leu Glu Glu Phe Asp Ala Pro Pro Glu Ala Val Glu Ala Asn Leu
50      55      60
Lys Val Lys Arg Cys Ile Asn Lys Ile Met Tyr Gly Asp Arg Leu Ser
65      70      75      80
Met Gly Thr Ser Leu Val Phe Ile Met Leu Lys Cys Asp Val Lys Val
85      90      95
Trp Leu Gln Ile Asn Phe Pro Arg Gly Arg Trp Phe Ser Glu Ile Asn
100      105      110

```

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 91 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```

Met Lys Leu Ala Ile Thr Leu Ala Leu Val Thr Leu Ala Leu Leu Cys
1      5      10      15

```

Ser Pro Ala Ser Ala Gly Ile Cys Pro Arg Phe Ala His Val Ile Glu  
 20 25 30  
 Asn Leu Leu Leu Gly Thr Pro Ser Ser Tyr Glu Thr Ser Leu Lys Glu  
 35 40 45  
 Phe Glu Pro Asp Asp Thr Met Lys Asp Ala Gly Met Gln Met Lys Lys  
 50 55 60  
 Val Leu Asp Ser Leu Pro Gln Thr Thr Arg Glu Asn Ile Met Lys Leu  
 65 70 75 80  
 Thr Glu Lys Ile Val Lys Ser Pro Leu Cys Met  
 85 90

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 93 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Lys Leu Leu Met Val Leu Met Leu Ala Ala Leu Ser Gln His Cys  
 1 5 10 15  
 Tyr Ala Gly Ser Gly Cys Pro Leu Leu Glu Asn Val Ile Ser Lys Thr  
 20 25 30  
 Ile Asn Pro Gln Val Ser Lys Thr Glu Tyr Lys Glu Leu Leu Gln Glu  
 35 40 45  
 Phe Ile Asp Asp Asn Ala Thr Thr Asn Ala Ile Asp Glu Leu Lys Glu  
 50 55 60  
 Cys Phe Leu Asn Gln Thr Asp Glu Thr Leu Ser Asn Val Glu Val Phe  
 65 70 75 80  
 Met Gln Leu Ile Tyr Asp Ser Ser Leu Cys Asp Leu Phe  
 85 90

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 503 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GACAGCGGCT TCCTTGATCC TTGCCACCCG CGACTGACACA CCGACAGCAG CAGCCTCACC 60

ATG AAG TTG CTG ATG GTC CTC ATG CTG GCG GCC CTC TCC CAG CAC TGC 108  
Met Lys Leu Leu Met Val Leu Met Leu Ala Ala Leu Ser Gln His Cys  
1 5 10 15

TAC GCA GGC TCT GGC TGC CCC TTA TTG GAG AAT GTG ATT TCC AAG ACA 156  
Tyr Ala Gly Ser Gly Cys Pro Leu Leu Glu Asn Val Ile Ser Lys Thr  
20 25 30

ATC AAT CCA CAA GTG TCT AAG ACT GAA TAC AAA GAA CTT CTT CAA GAG 204  
Ile Asn Pro Gln Val Ser Lys Thr Glu Tyr Lys Glu Leu Leu Gln Glu  
35 40 45

TTC ATA GAC GAC AAT GCC ACT ACA AAT GCC ATA GAT GAA TTG AAG GAA 252  
Phe Ile Asp Asp Asn Ala Thr Thr Asn Ala Ile Asp Glu Leu Lys Glu  
50 55 60

TGT TTT CTT AAC CAA ACG GAT GAA ACT CTG AGC AAT GTT GAG GTG TTT 300  
Cys Phe Leu Asn Gln Thr Asp Glu Thr Leu Ser Asn Val Glu Val Phe  
65 70 75 80

ATG CAA TTA ATA TAT GAC AGC AGT CTT TGT GAT TTA TTT TAA CTT TCT 348  
Met Gln Leu Ile Tyr Asp Ser Ser Leu Cys Asp Leu Phe  
85 90

GCA AGA CCT TTG GCT CAC AGA ACT GCA GGG TAT GGT GAG AAA CCA ACT 396

ACG GAT TGC TGC AAA CCA CAC CTT CTC TTT CTT ATG TCT TTT TAC TAC 444

AAA CTA CAA GAC AAT TGT TGA AAC CTG CTA TAC ATG TTT ATT TTA ATA 492

AAT TGA TGG CA 503

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 95 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Lys Leu Val Phe Leu Phe Leu Leu Val Thr Ile Pro Ile Cys Cys  
1 5 10 15

Tyr Ala Ser Gly Ser Gly Cys Ser Ile Leu Asp Glu Val Ile Arg Gly  
20 25 30

Thr Ile Asn Ser Thr Val Thr Leu His Asp Tyr Met Lys Leu Val Lys  
35 40 45

Pro Tyr Val Gln Asp His Phe Thr Glu Lys Ala Val Lys Gln Phe Lys  
50 55 60

Gln Cys Phe Leu Asp Gln Thr Asp Lys Thr Leu Glu Asn Val Gly Val  
65 70 75 80

Met Met Glu Ala Ile Phe Asn Ser Glu Ser Cys Gln Gln Pro Ser  
35 90 95

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 509 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGTTTCCTGA TTTCTGTCTT GGACAACAGA ACAACCCACA GGGACTGCCT CAAC ATG	57
Met	
1	
AAG CTG GTG TTT CTA TTC TTG TTG GTC ACC ATC CCT ATT TGC TGC TAT	105
Lys Leu Val Phe Leu Phe Leu Leu Val Thr Ile Pro Ile Cys Cys Tyr	
5 10 15	
GCC AGT GGT TCT GGC TGC AGT ATT CTA GAT GAA GTT ATT AGA GGT ACA	153
Ala Ser Gly Ser Gly Cys Ser Ile Leu Asp Glu Val Ile Arg Gly Thr	
20 25 30	
ATT AAC TCA ACT GTG ACT TTA CAT GAC TAT ATG AAA TTA GTT AAG CCA	201
Ile Asn Ser Thr Val Thr Leu His Asp Tyr Met Lys Leu Val Lys Pro	
35 40 45	
TAT GTA CAA GAT CAT TTT ACT GAA AAG GCT GTG AAG CAA TTC AAG CAG	249
Tyr Val Gln Asp His Phe Thr Glu Lys Ala Val Lys Gln Phe Lys Gln	
50 55 60 65	
TGT TTT CTA GAT CAG ACC GAC AAG ACT CTG GAA AAT GTT GGC GTG ATG	297
Cys Phe Leu Asp Gln Thr Asp Lys Thr Leu Glu Asn Val Gly Val Met	
70 75 80	
ATG GAG GCA ATA TTT AAC AGT GAA AGC TGT CAA CAG CCA TCC TAA ACA	345
Met Glu Ala Ile Phe Asn Ser Glu Ser Cys Gln Gln Pro Ser	
85 90 95	
TCT ACA AGA TCT TTG GCC ACA GGA CTC CAG GAA ACT GGC AAT GGC CAA	393
GCA ACT GAT AAC ACA GAT CAT AAC TCT TCT TTC TTG AAC CCC TTT TTC	441
TAC CTA TAA AGT GCA AGA CGA TTG TTG AAA CCT CAA ATT TAT GTC TTT	489
CCA TTT TAT TTA ATT ATC TG	509

## CLAIMS

1. A substantially purified human steroid binding protein C1 (hSBP1) polypeptide comprising the amino acid sequence of SEQ ID NO:1 or fragments thereof.
2. An isolated and purified polynucleotide sequence encoding an hSBP1 polypeptide of claim 1.
3. An isolated and purified polynucleotide sequence of claim 2 consisting of SEQ ID NO:2 or variants thereof.
4. A polynucleotide sequence which is complementary to SEQ ID NO:2 or degenerate variants thereof.
5. A recombinant expression vector comprising the polynucleotide sequence of claim 2.
6. A recombinant host cell containing the polynucleotide sequence of claim 5.
7. A method for producing a polypeptide comprising a polypeptide of SEQ ID NO:1, the method comprising the steps of:
  - a) culturing the host cell of claim 6 under conditions suitable for the expression of the polypeptide; and
  - b) recovering the polypeptide from the host cell culture.
8. A pharmaceutical composition comprising a substantially purified hSBP polypeptide having an amino acid sequence of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.
9. A purified antibody that specifically binds the polypeptide of claim 1.
10. A purified antagonist which specifically regulates or modulates the activity of the polypeptide of claim 1.
11. A pharmaceutical composition comprising a substantially purified antagonist of the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.
12. A substantially purified human steroid binding protein C2 (hSBP2) polypeptide comprising the amino acid sequence of SEQ ID NO:3 or fragments thereof.
13. An isolated and purified polynucleotide sequence encoding an hSBP2 polypeptide of claim 12.
14. An isolated and purified polynucleotide sequence of claim 13 consisting of SEQ ID NO:4 or variants thereof.
15. A polynucleotide sequence which is complementary to SEQ ID NO:4 or degenerate variants thereof.

16. A recombinant expression vector comprising the polynucleotide sequence of claim 13.
17. A recombinant host cell containing the polynucleotide sequence of claim 13.
18. A method for producing a polypeptide comprising a polypeptide of SEQ ID NO:3, the method comprising the steps of:
  - 5 a) culturing the host cell of claim 17 under conditions suitable for the expression of the polypeptide; and
  - b) recovering the polypeptide from the host cell culture.
19. A pharmaceutical composition comprising a substantially purified human steroid binding protein C2 (hSBP2) polypeptide having an amino acid sequence of SEQ ID NO:3 in conjunction  
10 with a suitable pharmaceutical carrier.
20. A purified antibody that specifically binds the polypeptide of claim 12.
21. A purified antagonist which specifically regulates or modulates the activity of the polypeptide of claim 12.
22. A pharmaceutical composition comprising a substantially purified antagonist of the  
15 polypeptide of claim 12 in conjunction with a suitable pharmaceutical carrier.

1/11

```

          9      18      27      36      45      54
5' NGT CCA AAT CAC TCA TTG TTT GTG AAA GCT GAG CTC ACA GCA AAA CAA GCC ACC

          63      72      81      90      99      108
ATG AAG CTG TCG GTG TGT CTC CTG CTG GTC ACG CTG GCC CTC TGC TGC TAC CAG
M K L S V C L L L V T L A L C C Y Q

          117      126      135      144      153      162
GCC AAT GCC GAG TTC TGC CCA GCT CTT GTT TCT GAG CTG TTA GAC TTC TTC TTC
A N A E F C C P A L V S E L L D F F

          171      180      189      198      207      216
ATT AGT GAA CCT CTG TTC AAG TTA AGT CTT GCC AAA TTT GAT GCC CCT CCG GAA
I S E P L F K L S L A K F D A P P E

          225      234      243      252      261      270
GCT GTT GCA GCC AAG TTA GGA GTG AAG AGA TGC ACG GAT CAG ATG TCC CTT CAG
A V A A K L G V K R C T D Q M S L Q

          279      288      297      306      315      324
AAA CGA AGC CTC ATT GCG GAA GTC CTG GTG AAA ATA TTG AAG AAA TGT AGT GTG
K R S L I A E V L V K I L K K C S V

          333      342      351      360      369      378
TGA CAT GTA AAA ACT TTC ATC CTG GTT TCC ACT GTC TTT CAA TGA CAC CCT GAT

          387      396      405
CTT CAC TGC AGA ATG TAA AGG TTT CAA C 3'

```

FIGURE 1

	9	18	27	36	45	54
5' GAT CCT TGC CAC CCG CGA CTG AAC ACC GAC AGC AGC AGC CTC ACC ATG AAG TTG	M K L					
	63	72	81	90	99	108
CTG ATG GTC CTC ATG CTG GCG GCC CTC TCC CAG CAC TGC TAC GCA GGC TCT TGT	L M V L L M L A A L S Q H C Y A G S G					
	117	126	135	144	153	162
TGC CCC TTA TTG GAG AAT GTG ATT TCC AAG ACA ATC AAT CCA CAA GTG TCT AAG	C P L L E N V I S K T I N P Q V S K					
	171	180	189	198	207	216
ACT GAA TAC AAA GAA CTT CTT CAA GAG TTC ATA GAC GAC AAT GCC ACT ACA AAT	T E Y K E L L L Q E F I D D N A T T N					
	225	234	243	252	261	270
GCC ATA GAT GAA TTG AAG GAA TGT TTT CTT AAC CAA ACG GAT GAA ACT CTG AGC	A I D E E L K E C F L N Q T D E T L S					
	279	288	297	306	315	324
AAT GTT GAG GTG TTT ATG CAA TTA ATA TAT GAC AGC AGT CTT TGT GAT TTA TTT	N V E V F M Q Q L I Y D S S L C D L F					
	333	342	351	360	369	378
TAA CTT TCT GCA AGA CCT TTG GCT CAC AGA ACT GCA GGG TAT GGT GAG AAA CCA						

FIGURE 2A



3/11

387                    396                    405                    414                    423                    432  
ACT ACG GAT TGC TGC AAA CCA CAC CTT CTC TTT CTT ATG TCT TTT TAC TAC AAA

441                    450                    459                    468                    477                    486  
CTA CAA GAC AAT TGT TGA AAC CTG CTA TAC ATG TTT ATT TTA ATA AAT TGA TGG

495  
CAA AAA CTG 3'

FIGURE 2B

4/11

Library	Lib Description	Abun	Pct Abun
BRSTTUT01	breast tumor, 55 F, match to BRSTNOT02	12	0.1811
THYMNOT02	thymus, 3 M	2	0.0386
BRSTNOT02	breast, 55 F, match to BRSTTUT01	1	0.0192
BRSTNOT01	breast, 56 F	1	0.0158
PROSNOT06	prostate, 57 M, match to PROSTUT04	1	0.0154
BRSTNOT05	breast, 58 F, match to BRSTTUT03	1	0.0148

FIGURE 3

5/11

Library	Lib Description	Abun	Pct Abun
BRSTTUT03	breast tumor, 58 F, match to BRSTNOT05	28	0.4149
BRSTTUT01	breast tumor, 55 F, match to BRSTNOT02	26	0.3923
BRSTNOT02	breast, 55 F, match to BRSTTUT01	5	0.0792
BRSTNOT05	breast, 58 F, match to BRSTTUT03	4	0.0617
BRSTNOT03	breast, 54 F, match to BRSTTUT03	3	0.0440
BRSTNOT01	breast, 56 F	2	0.0384
BRSTNOT04	breast, 62 F	1	0.0319

FIGURE 4

6/11

1	M	-	-	-	K	L	S	V	C	L	L	L	V	T	L	A	L	C	C	Y	Q	A	N	A	E	-	-	-	-	F	C	P	A	L	V	S	E	L	L	SEQ ID NO:1	
1	M	S	T	I	K	L	S	L	C	L	L	I	M	-	L	A	V	C	C	Y	E	A	N	-	-	-	-	A	S	Q	I	C	E	L	V	A	H	E	T	I	SEQ ID NO:5
1	M	-	-	-	R	L	S	L	C	L	L	T	I	-	L	V	V	C	C	Y	E	A	N	G	Q	T	L	A	G	Q	V	C	Q	A	L	Q	D	V	T	I	SEQ ID NO:8
1	M	-	-	-	K	L	A	I	T	L	A	L	V	T	L	A	L	C	S	P	A	S	A	G	-	-	-	-	I	C	P	R	F	A	H	V	I	E	SEQ ID NO:9		
33	D	F	F	F	I	S	E	P	L	F	K	L	S	L	A	K	F	D	A	P	P	E	A	V	A	A	K	L	G	V	K	R	C	T	D	Q	M	S	L	Q	SEQ ID NO:1
36	S	F	L	M	K	S	E	E	E	L	K	K	E	L	E	M	Y	N	A	P	P	A	V	E	A	A	K	L	E	V	K	R	C	V	D	Q	M	S	N	G	SEQ ID NO:5
37	T	F	L	N	P	E	E	E	E	L	K	R	E	L	E	E	F	D	A	P	P	E	A	V	E	A	N	L	K	V	K	R	C	I	N	K	I	M	Y	G	SEQ ID NO:8
33	N	L	L	G	T	P	S	S	Y	E	T	S	L	K	E	F	E	P	D	D	T	M	K	D	A	G	M	Q	M	K	K	V	L	D	S	L	P	Q	T	SEQ ID NO:9	
73	K	R	S	L	I	A	E	V	L	V	K	I	L	K	K	C	S	V																			SEQ ID NO:1				
76	D	R	L	V	V	A	E	T	L	V	Y	I	F	L	E	C	G	V	K	Q	W	V	E	T	Y	Y	P	E	I	D	F	Y	Y	D	M	N			SEQ ID NO:5		
77	D	R	L	S	M	G	T	S	L	V	F	I	M	L	K	C	D	V	K	V	W	L	Q	I	N	F	P	R	G	R	W	F	S	E	I	N			SEQ ID NO:8		
73	T	R	E	N	I	M	K	-	L	T	E	K	I	V	K	S	P	L	C	M																	SEQ ID NO:9				

FIGURE 5

7/11

1	M	K	L	L	M	V	L	M	L	A	A	L	S	Q	H	C	Y	A	G	S	G	C	P	L	L	E	N	V	I	S	K	T	I	N	P	Q	V	S	K	T	SEQ ID NO:3
1	M	K	L	L	M	V	L	M	L	A	A	L	S	Q	H	C	Y	A	G	S	G	C	P	L	L	E	N	V	I	S	K	T	I	N	P	Q	V	S	K	T	SEQ ID NO:10
1	M	K	L	V	F	L	F	L	L	V	T	I	P	I	C	C	Y	A	S	G	S	G	C	S	I	L	D	E	V	I	R	G	T	I	N	S	T	V	T	L	SEQ ID NO:12
41	E	Y	K	E	L	L	Q	E	F	I	D	D	N	A	T	T	N	A	I	D	E	L	K	E	C	F	L	N	Q	T	D	E	T	L	S	N	V	E	V	F	SEQ ID NO:3
41	E	Y	K	E	L	L	Q	E	F	I	D	D	N	A	T	T	N	A	I	D	E	L	K	E	C	F	L	N	Q	T	D	E	T	L	S	N	V	E	V	F	SEQ ID NO:10
41	H	D	Y	M	K	L	V	K	P	Y	V	Q	D	H	F	T	E	K	A	V	K	Q	F	K	C	F	L	D	Q	T	D	K	T	L	E	N	V	G	V	SEQ ID NO:12	
81	M	Q	L	I	Y	D	S	S	S	L	C	D	L	F																										SEQ ID NO:3	
81	M	Q	L	I	Y	D	S	S	S	L	C	D	L	F																									SEQ ID NO:10		
81	M	M	E	A	I	F	N	S	E	S	C	Q	Q	P	S																								SEQ ID NO:12		

FIGURE 6



9/11

238	T	G	A	A	G	A	T	G	C	A	C	G	A	T	C	A	G	A	A	C	G	A	SEQ	ID NO:2
226	T	G	A	A	G	A	A	T	G	T	A	T	A	G	A	C	A	G	A	A	G	A	SEQ	ID NO:7
213	T	G	A	A	G	A	T	G	T	G	T	A	G	A	C	C	A	G	A	A	T	G	SEQ	ID NO:6
276	A	G	C	C	T	C	A	T	T	-	G	C	G	A	A	G	T	C	T	G	G	T	SEQ	ID NO:2
266	A	A	T	G	G	A	A	A	T	G	G	A	A	A	T	A	C	T	G	G	C	A	SEQ	ID NO:7
250	A	T	T	G	G	T	A	G	C	A	G	A	A	A	C	A	C	T	G	T	A	T	SEQ	ID NO:6
315	T	G	T	A	G	T	G	T	G	A	C	A	T	-	-	G	T	A	A	A	C	T	SEQ	ID NO:2
306	T	G	T	A	A	A	G	-	-	G	A	A	C	A	T	G	A	A	A	C	A	T	SEQ	ID NO:7
290	T	G	T	G	T	G	A	A	A	C	A	A	T	G	G	T	A	G	A	A	C	A	SEQ	ID NO:6
349	T	T	C	C	A	C	T	G	T	C	T	T	C	A	T	T	C	A	A	T	C	T	SEQ	ID NO:2
344	T	C	T	C	C	A	T	T	G	C	T	T	C	T	C	A	C	A	G	A	T	A	SEQ	ID NO:7
328	G	A	T	C	G	A	T	T	C	T	A	C	G	A	T	A	T	G	A	A	C	T	SEQ	ID NO:6
389	A	A	T	G	T	A	A	G	G	T	T	T	C	A	A	C							SEQ	ID NO:2
384	A	A	T	G	T	G	A	A	G	G	T	T	T	C	A	A	C	G	T	C	T	T	SEQ	ID NO:7
368	A	A	T	G	T	G	A	T	G	G	T	T	T	C	A	A	-	G	T	C	T	T	SEQ	ID NO:6
405																							SEQ	ID NO:2
424	C	C	T	T	G	C	A	T	G	T	T	A	A	A	A	A	A						SEQ	ID NO:7
407	C	C	T	T	G	C																SEQ	ID NO:6	

FIGURE 7B

10/11

```

1  GA - - - - - T C C T T G C C A C C C G C G A C T G A A C A SEQ ID NO:4
1  G A C A G C G G C T T C T T G A T C C C T T G C C A C C C G C G A C T G A A C A SEQ ID NO:11
1  A G - - - - - T T T C C T G A T T T C T G T C T T G G A C A C A G A A C A SEQ ID NO:13

26  - C C G A C A G C A G C C C T C A C C A T G A A G T T G C T G A T G G T C C SEQ ID NO:4
41  - C C G A C A G C A G C C C T C A C C A T G A A G T T G C T G A T G G T C C SEQ ID NO:11
34  A C C C A C A G G G A C T G C C C T C A C A T G A A G C T G T G T T C T A T SEQ ID NO:13

65  T C A T G C T G G C C G C C C T C T C C C A G C A C T G C T A C G C - A G - - G SEQ ID NO:4
80  T C A T G C T G G C C G C C C T C T C C C A G C A C T G C T A C G C - A G - - G SEQ ID NO:11
74  T C T T G T T G G T C A C C A T C C T A T T T G C T G C T A T G C C A G T G G SEQ ID NO:13

102 C T C T G G C T G C C C C C T T A T T G G A G A A T G T G A T T T C C A A G A C A SEQ ID NO:4
117 C T C T G G C T G C C C C C T T A T T G G A G A A T G T G A T T T C C A A G A C A SEQ ID NO:11
114 T T C T G G C T G C A G T A T T C T A G A T G A A G T A T T A G A G G T A C A SEQ ID NO:13

142 A T C A A T C C A C A A G T G T C T A A G A C T G A A T A C A A A G A A C T T C SEQ ID NO:4
157 A T C A A T C C A C A A G T G T C T A A G A C T G A A T A C A A A G A A C T T C SEQ ID NO:11
154 A T T A A C T C A A C T G T G A C T T T A C A T G A C T A T A T G A A A T T A G SEQ ID NO:13

182 T T C A A G A G T T C A T A G A C G A C A A T G C C A C T A C A A A T G C C A T SEQ ID NO:4
197 T T C A A G A G T T C A T A G A C G A C A A T G C C A C T A C A A A T G C C A T SEQ ID NO:11
194 T T A A G C C A T A T G T A C A A G A T C A T T T A C T G A A A A G G C T G T SEQ ID NO:13

222 A G A T G A A T T G A A G G A A T G T T T T C T T A A C C A A A C G G A T G A A SEQ ID NO:4
237 A G A T G A A T T G A A G G A A T G T T T T C T T A A C C A A A C G G A T G A A SEQ ID NO:11
234 G A A G C A A T T C A A G C A G T G T T T T C T T A G A T C A G A C C G A C A A G SEQ ID NO:13

```

FIGURE 8A



11/11

262 A C T C T G A G C A A T G T T T G A G G T G T T T A T G C A A T T A A T A T A T G S E Q I D N O : 4  
277 A C T C T G A G C A A T G T T T G A G G T G T T T A T G C A A T T A A T A T G S E Q I D N O : 11  
274 A C T C T G G A A A A T G T T G G C G T G A T G A T G G C A A T A T T A S E Q I D N O : 13

302 A C A G C - - - A G T C T T T G T G A T T T A A C T T T C T G C A A G S E Q I D N O : 4  
317 A C A G C - - - A G T C T T T G T G A T T T A A C T T T C T G C A A G S E Q I D N O : 11  
314 A C A G T G A A A G C T G T C A A C A G C C A T C C T A A A C A T C T A C A G S E Q I D N O : 13

339 A C C T T T G G C T C A C A G A A C T G C A G G G T A - T G G T G A G A A A C C S E Q I D N O : 4  
354 A C C T T T G G C T C A C A G A A C T G C A G G G T A - T G G T G A G A A A C C S E Q I D N O : 11  
354 A T C T T T G G C - C A C A G G A C T C C A G G A A A C T G G C A A T G G C C A S E Q I D N O : 13

378 A A C T A C G G A T T G C T G C A A A C C A C A C C T T C T C T T C T T A T G S E Q I D N O : 4  
393 A A C T A C G G A T T G C T G C A A A C C C A C A C C T T C T C T T A T G S E Q I D N O : 11  
393 A G C A C T G A T A A C - A C A G A T C A T A C T C T T C T T G A A S E Q I D N O : 13

418 T C - - - T T T T A C - T A C A A A C T A C A A G A C A A T T G T T G A A A S E Q I D N O : 4  
433 T C - - - T T T T A C - T A C A A A C T A C A A G A C A A T T G T T G A A A S E Q I D N O : 11  
432 C C C C T T T T C T A C C T A T A A A G T G C A A G A C G A T T G T T G A A A S E Q I D N O : 13

453 C C T G C T A T A C A T G T T T A T T T A A T A A A T T G A T G G C A A A S E Q I D N O : 4  
468 C C T G C T A T A C A T G T T T A T T T T A A T A A A T T G A T G G C A S E Q I D N O : 11  
472 C C T C A A A T T T A T G T C T - T T C C A T T T A T T A - - - A T T A T S E Q I D N O : 13

493 C T G S E Q I D N O : 4  
503 S E Q I D N O : 11  
507 C T G S E Q I D N O : 13

FIGURE 8B

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/20674

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C07K14/47 C07K16/18 C12Q1/68 G01N33/574  
G01N33/68 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K C12N C12Q G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WATSON M A ET AL: "MAMMAGLOBIN, A MAMMARY-SPECIFIC MEMBER OF THE UTEROGLOBIN GENE FAMILY, IS OVEREXPRESSED IN HUMAN BREAST CANCER" CANCER RESEARCH, vol. 56, 15 February 1996, pages 860-865, XP002048615 cited in the application see abstract; figures 1-4 see page 862, paragraph 2 - page 864, paragraph 1	12-18,20
A		19,21,22

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

16 March 1998

Date of mailing of the international search report

26/03/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040. Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer:

Oderwald, H

# INTERNATIONAL SEARCH REPORT

Inter. onal Application No

PCT/US 97/20674

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	<p>WO 96 38463 A (UNIV WASHINGTON ; WATSON MARK A (US); FLEMING TIMOTHY P (US)) 5 December 1996 see the claims see abstract; figures 2, 4A; examples 1, 3, 5 see page 3, line 18 - page 5, line 17 see page 19, line 6 - page 21, line 4</p> <p>---</p>	12-18, 20
A	<p>WATSON M A ET AL: "ISOLATION OF DIFFERENTIALLY EXPRESSED SEQUENCE TAGS FROM HUMAN BREAST CANCER" CANCER RESEARCH, vol. 54, 1 September 1994, pages 4598-4602, XP002019977 see abstract; figures 2, 4 see page 4599, paragraph 3 - paragraph 4</p> <p>---</p>	
A	<p>PARKER M ET AL: "PROSTATIC STEROID BINDING PROTEIN: GENE DUPLICATION AND STEROID BINDING" NATURE, vol. 298, no. 5869, 1 July 1982, pages 92-94, XP002048614 see abstract; figures 1, 2 see page 93, paragraph 3 - page 94, paragraph 3</p> <p>---</p>	1-11
A	<p>PARKER M G ET AL.: "Prostatic steroid-binding protein" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 258, no. 1, 10 January 1983, pages 12-15, XP002058665 cited in the application see the whole document</p> <p>-----</p>	12-22

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 97/20674

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9638463 A	05-12-96	US 5668267 A	16-09-97
		AU 5961696 A	18-12-96
<hr/>			